

DIPARTIMENTO DI SCIENZE E TECNOLOGIE  
BIOLOGICHE CHIMICHE E FARMACEUTICHE (STEBICEF)

Piano Lauree Scientifiche

# Scienze omiche: Grandi apparecchiature per la Proteomica

Mercoledì 27 Luglio 2022

Patrizia Cancemi

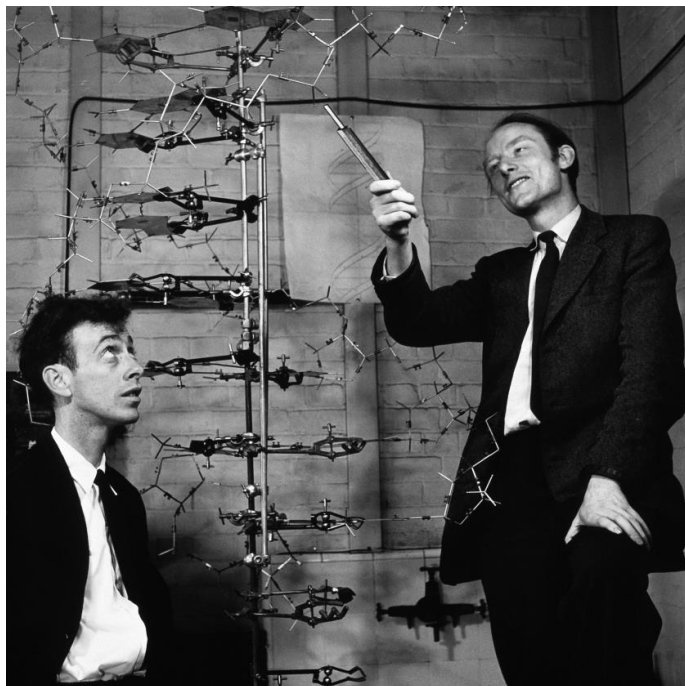


# Di che cosa parleremo?

1. Perché si sono sviluppate le scienze omiche
2. Quali sono le scienze omiche
3. Proteomica e tecnologie di indagine high-throughput
4. La spettrometria di massa per identificare le proteine
5. Integrazione delle scienze omiche
6. L'approccio delle scienze omiche con gli studenti

# Dal dogma centrale della biologia molecolare alle scienze omiche

Scienze omiche: Grandi apparecchiature per la Proteomica



No. 4356 April 25, 1953 NATURE 737

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.H.S. *Discovery II* for their part in making the observations.

- <sup>1</sup> Young, F. B., Gerard, H., and Jernon, W., *Phil. Mag.*, **40**, 149 (1925).
- <sup>2</sup> Longuet-Higgins, M. S., *Mon. Not. Roy. Astr. Soc., Geophys. Supp.*, **8**, 255 (1949).
- <sup>3</sup> Von Arx, W. S., *Woods Hole Papers in Phys., Oceanogr., Meteor.*, **11** (5) (1946).
- <sup>4</sup> Ekman, V. W., *Arkiv. Mat. Astron. Fysik.* (Stockholm), **2** (11) (1935).

## MOLECULAR STRUCTURE OF NUCLEIC ACIDS

### A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey<sup>1</sup>. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining  $\beta$ -D-deoxyribofuranose residues with 3',5' linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's<sup>2</sup> model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There



This figure is purely diagrammatic. The two ribbons symbolize the two phosphate-sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis.

is a residue on each chain every 3.4 Å. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 Å. The distance of a phosphorus atom from the fibre axis is 10 Å. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-coordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally<sup>3,4</sup> that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyriboses, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data<sup>3,4</sup> on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

738 NATURE April 25, 1953 Vol. 171

King's College, London. One of us (J.D.W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. WATSON  
F. H. C. CRICK

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge, April 2.

- <sup>1</sup> Pauling, L., and Corey, R. B., *Nature*, **171**, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, **32**, 84 (1953).
- <sup>2</sup> Furberg, S., *Acta Chem. Scand.*, **6**, 634 (1952).
- <sup>3</sup> Chargaff, E., for references see Zamenhof, S., Braverman, G., and Chargaff, E., *Biochim. et Biophys. Acta*, **9**, 612 (1952).
- <sup>4</sup> Wyatt, G. H., *J. Gen. Physiol.*, **26**, 201 (1952).
- <sup>5</sup> Astbury, W. T., *Synop. Soc. Exp. Biol.*, **1**, *Nucleic Acid*, 66 (Camb. Univ. Press, 1947).
- <sup>6</sup> Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, **10**, 192 (1953).

### Molecular Structure of Deoxyribose Nucleic Acids

WHILE the biological properties of deoxyribose nucleic acid suggest a molecular structure containing great complexity, X-ray diffraction studies described here (cf. Astbury<sup>1</sup>) show the basic molecular configuration has great simplicity. The purpose of this communication is to describe, in a preliminary way, some of the experimental evidence for the polynucleotide chain configuration being helical, and existing in this form when in the natural state. A fuller account of the work will be published shortly.

The structure of deoxyribose nucleic acid is the same in all species (although the nitrogen base ratios alter considerably) in nucleoprotein, extracted or in cells, and in purified nucleate. The same linear group of polynucleotide chains may pack together parallel in different ways to give crystalline<sup>2-4</sup>, semi-crystalline or paracrystalline material. In all cases the X-ray diffraction photograph consists of two regions, one determined largely by the regular spacing of nucleotides along the chain, and the other by the longer spacings of the chain configuration. The sequence of different nitrogen bases along the chain is not made visible.

Oriented paracrystalline deoxyribose nucleic acid (structure B<sup>5</sup> in the following communication by Franklin and Gosling) gives a fibre diagram as shown in Fig. 1 (of ref. 4). Astbury suggested that the strong 3.4-Å. reflexion corresponded to the inter-nucleotide repeat along the fibre axis. The ~34 Å. layer lines, however, are not due to a repeat of a polynucleotide composition, but to the chain configuration repeat, which causes strong diffraction as the nucleotide chains have higher density than the interstitial water. The absence of reflexions on or near the meridian immediately suggests a helical structure with axis parallel to fibre length.

#### Diffraction by Helices

It may be shown<sup>6</sup> (also Stokes, unpublished) that the intensity distribution in the diffraction pattern of a series of points equally spaced along a helix is given by the squares of Bessel functions. A uniform continuous helix gives a series of layer lines of spacing corresponding to the helix pitch, the intensity distribution along the *n*th layer line being proportional to the square of  $J_n$ , the *n*th order Bessel function. A straight line may be drawn approximately through

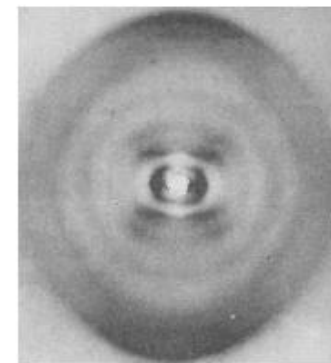


Fig. 1. Fibre diagram of deoxyribose nucleic acid from *R. coli*. Fibre axis vertical.

the innermost maxima of each Bessel function and the origin. The angle this line makes with the equator is roughly equal to the angle between an element of the helix and the helix axis. If a unit repeats *n* times along the helix there will be a meridional reflexion ( $J_0^2$ ) on the *n*th layer line. The helical configuration produces side-bands on this fundamental frequency, the effect<sup>6</sup> being to reproduce the intensity distribution about the origin around the new origin, on the *n*th layer line, corresponding to *C* in Fig. 2.

We will now briefly analyse in physical terms some of the effects of the shape and size of the repeat unit or nucleotide on the diffraction pattern. First, if the nucleotide consists of a unit having circular symmetry about an axis parallel to the helix axis, the whole diffraction pattern is modified by the form factor of the nucleotide. Second, if the nucleotide consists of a series of points on a radius at right-angles to the helix axis, the phases of radiation scattered by the helices of different diameter passing through each point are the same. Summation of the corresponding Bessel functions gives reinforcement for the inner-

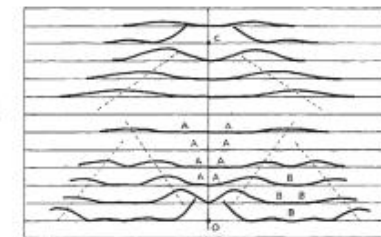
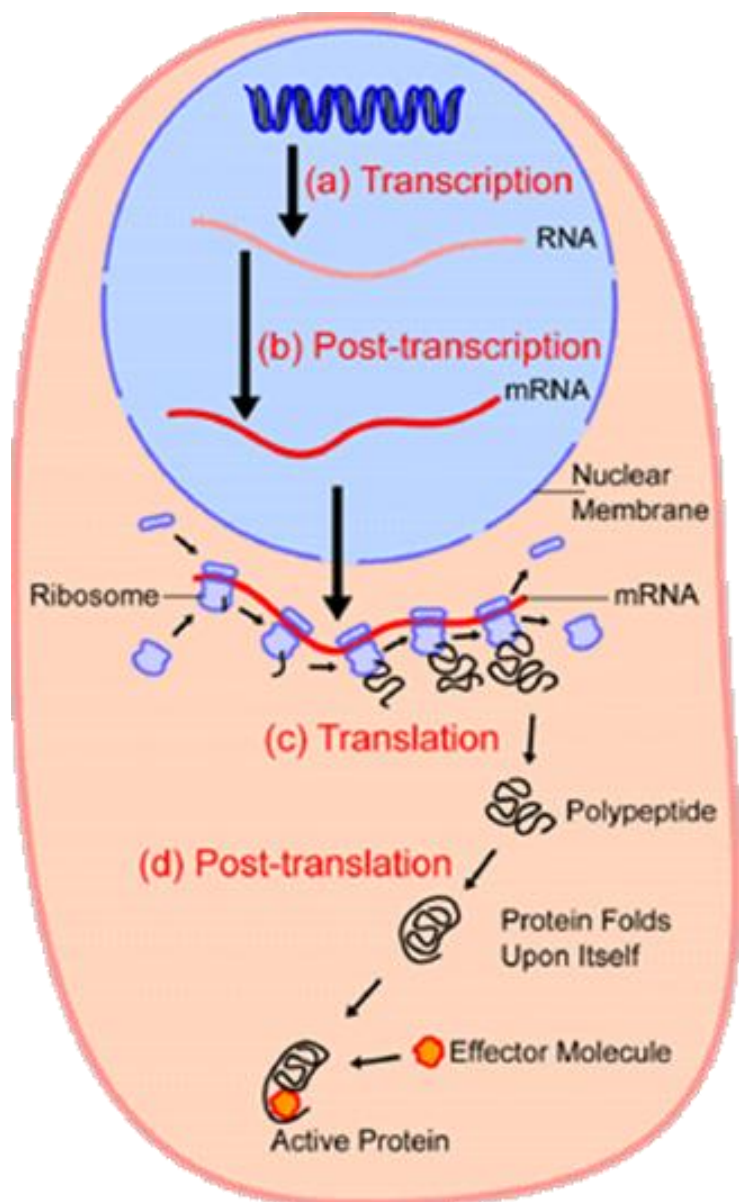


Fig. 2. Diffraction pattern of system of helices corresponding to structure of deoxyribose nucleic acid. The squares of Bessel functions are plotted about 0 on the equator and on the first, second, third and fifth layer lines for half of the nucleotide axes at 20 Å. diameter and remainder distributed along a radius, the axes at a given radius being proportional to the radius. About  $C$  on the tenth layer line similar functions are plotted for an inner diameter of 12 Å.

# Dal dogma centrale della biologia molecolare alle scienze omiche



**REPLICAZIONE**  
DNA  $\longrightarrow$  DNA

**TRASCRIZIONE**  
DNA  $\longrightarrow$  RNA

**TRADUZIONE**  
RNA  $\longrightarrow$  PROTEINE

L'informazione per la sintesi delle proteine è contenuta nel DNA. La trascrizione e la traduzione sono i processi attraverso cui l'informazione genetica è convertita in proteine. Il flusso dell'informazione genetica è unidirezionale, ovvero non passa mai dalle proteine agli acidi nucleici.

**Il codice genetico non è sovrapposto**  
**Il codice non ha punteggiatura**  
**Il codice genetico è degenerato**  
**Il codice genetico è letto in triplette**

# Dal dogma centrale della biologia molecolare alle scienze omiche

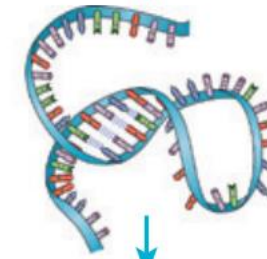
Scienze omiche: Grandi apparecchiature per la Proteomica

Epigenomica

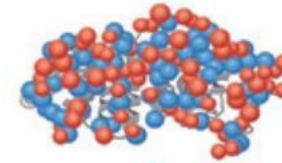
Genomica ← Geni



Trascrittomica ← mRNA



Proteomica ← Proteine



Metabolomica ← Metaboliti



Farmacogenomica

Nutrigenomica

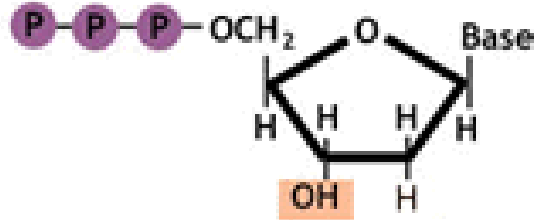
Microbioma

Lipidomica

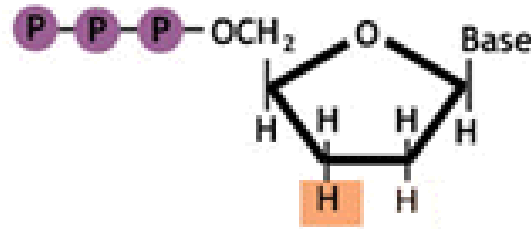
Glicomica

# Metodo di Sanger per il sequenziamento del DNA

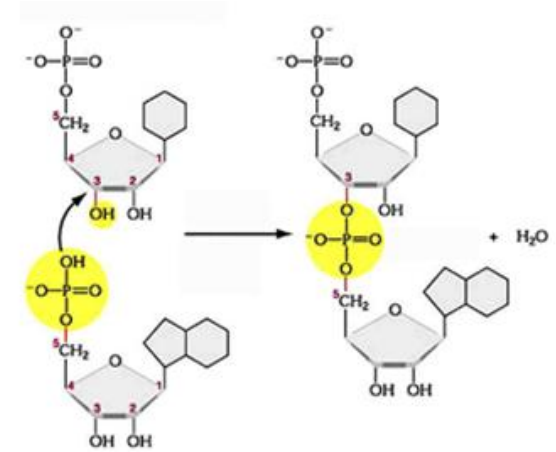
Metodo enzimatico (poiché sfrutta gli enzimi della replicazione del DNA)  
Basato sulla capacità di mimare in vitro la replicazione del DNA e sulla possibilità di interrompere la sintesi e identificare l'ultimo nucleotide inserito



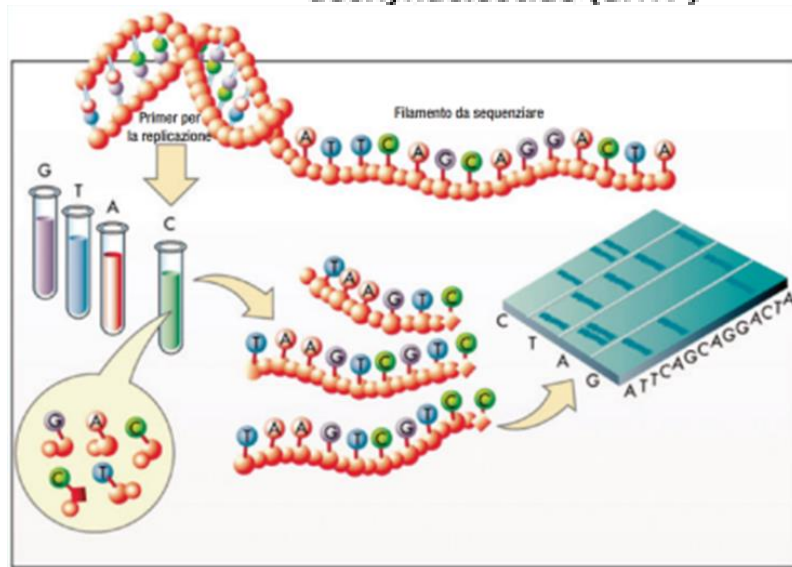
deoxynucleotide (dNTP)



dideoxynucleotide (ddNTP)



Formazione del legame tostodiesterico



Nature Vol. 265 February 24 1977

687

## articles

### Nucleotide sequence of bacteriophage $\Phi$ X174 DNA

F. Sanger, G. M. Air<sup>\*</sup>, B. G. Barrell, N. L. Brown<sup>1</sup>, A. R. Coulson, J. C. Fiddes, C. A. Hutchison III<sup>2</sup>, P. M. Slocombe<sup>3</sup> & M. Smith<sup>4</sup>

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

*A DNA sequence for the genome of bacteriophage  $\Phi$ X174 of approximately 5,375 nucleotides has been determined using the rapid and simple 'plus and minus' method. The sequence identifies many of the features responsible for the production of the proteins of the nine known genes of the organism, including initiation and termination sites for the proteins and RNAs. Two pairs of genes are coded by the same region of DNA using different reading frames.*

*The genome of bacteriophage  $\Phi$ X174 is a single-stranded, circular DNA of approximately 5,400 nucleotides coding for nine known proteins. The order of these genes, as determined by genetic techniques<sup>1-4</sup>, is A-B-C-D-E-J-F-G-H. Genes F, G and H code for structural proteins of the virus capsid, and gene*

*strand DNA of  $\Phi$ X has the same sequence as the mRNA and, in certain conditions, will bind ribosomes so that a protected fragment can be isolated and sequenced. Only one major site was found. By comparison with the amino acid sequence data it was found that this ribosome binding site sequence coded for the initiation of the gene G protein<sup>14</sup> (positions 2,362-2,413).*

*At this stage sequencing techniques using primed synthesis with DNA polymerase were being developed<sup>15</sup> and Schott<sup>16</sup> synthesised a decanucleotide with a sequence complementary to part of the ribosome binding site. This was used to prime into the intercistronic region between the F and G genes, using DNA polymerase and <sup>32</sup>P-labelled triphosphates<sup>15</sup>. The ribo-substitution technique<sup>15</sup> facilitated the sequence determination of the labelled DNA produced. This decanucleotide-primed system was also used to develop the plus and minus method<sup>1</sup>. Suitable synthetic primers are, however, difficult to prepare and as*

## The Nobel Prize in Chemistry 1980



Photo from the Nobel Foundation archive

Paul Berg

Prize share: 1/2



Photo from the Nobel Foundation archive

Walter Gilbert

Prize share: 1/4



Photo from the Nobel Foundation archive

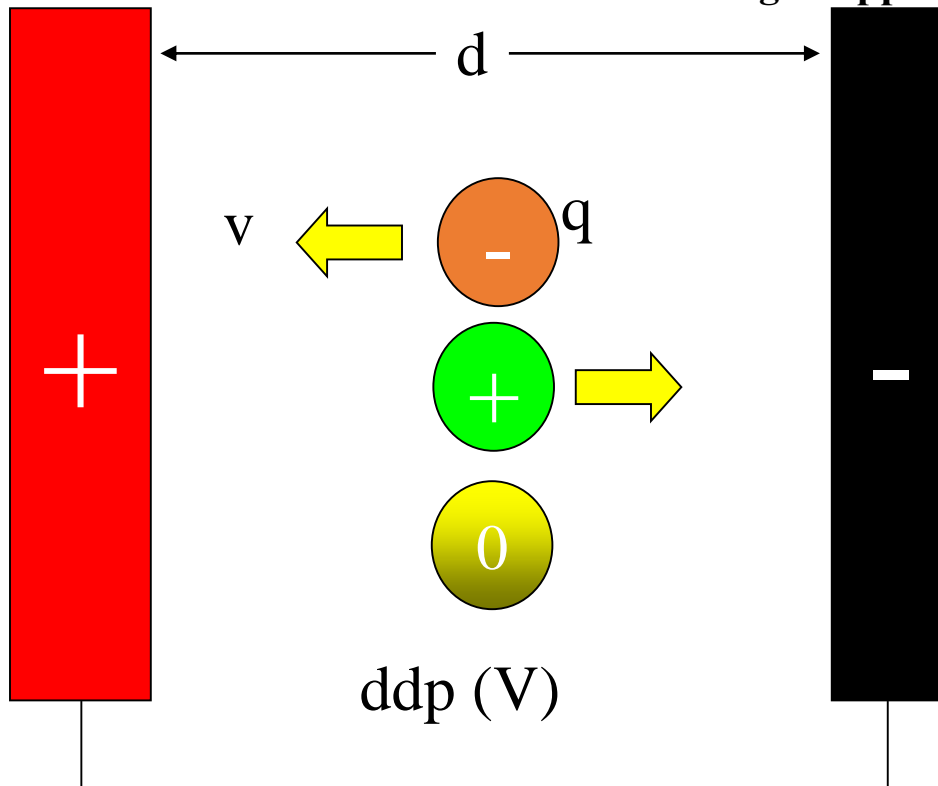
Frederick Sanger

Prize share: 1/4

The Nobel Prize in Chemistry 1980 was divided, one half awarded to Paul Berg "for his fundamental studies of the biochemistry of nucleic acids, with particular regard to recombinant-DNA", the other half jointly to Walter Gilbert and Frederick Sanger "for their contributions concerning the determination of base sequences in nucleic acids."

# Elettroforesi: principi generali

Una molecola con una carica netta non nulla posta tra due elettrodi di segno opposto migra verso l'elettrodo con segno opposto alla sua carica netta



**ddp (V):** differenza di potenziale applicata tra gli elettrodi

**d:** distanza tra gli elettrodi

**q:** carica della molecola

**E (campo elettrico) = V/d**

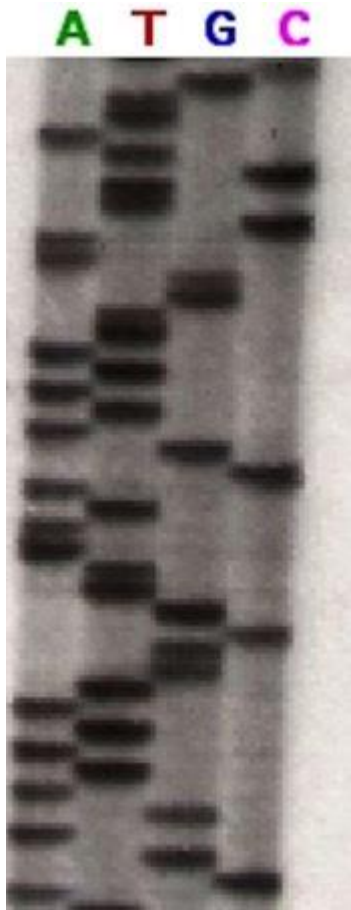
La molecola si muove verso l'elettrodo di segno opposto con una velocità ( $v$ ) che è proporzionale all'entità del campo elettrico ( $V$ ) e della sua carica ( $q$ ) e inversamente proporzionale all'attrito che incontra nel muoversi ( $f$  – coefficiente d'attrito)

# Elettroforesi su gel

- Metodo più utilizzato
- Separazione di proteine ed acidi nucleici

1. Il gel agisce da setaccio molecolare e migliora la separazione

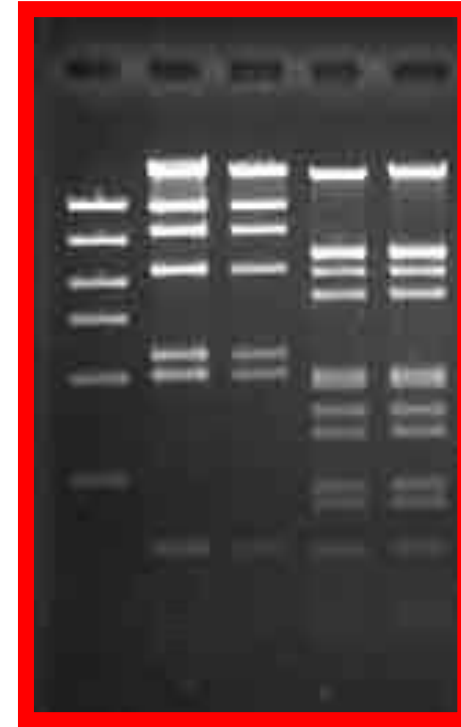
- Gel di agarosio
- Gel di poliacrilamide
- Gel di silice
- Gel di amido



L'elettroforesi può essere considerata una tecnica di separazione che ci permette di pesare il DNA

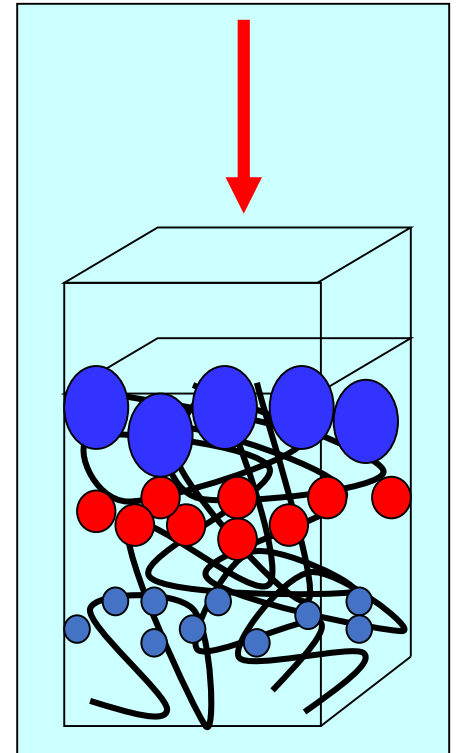


DNA PIU' PESANTE



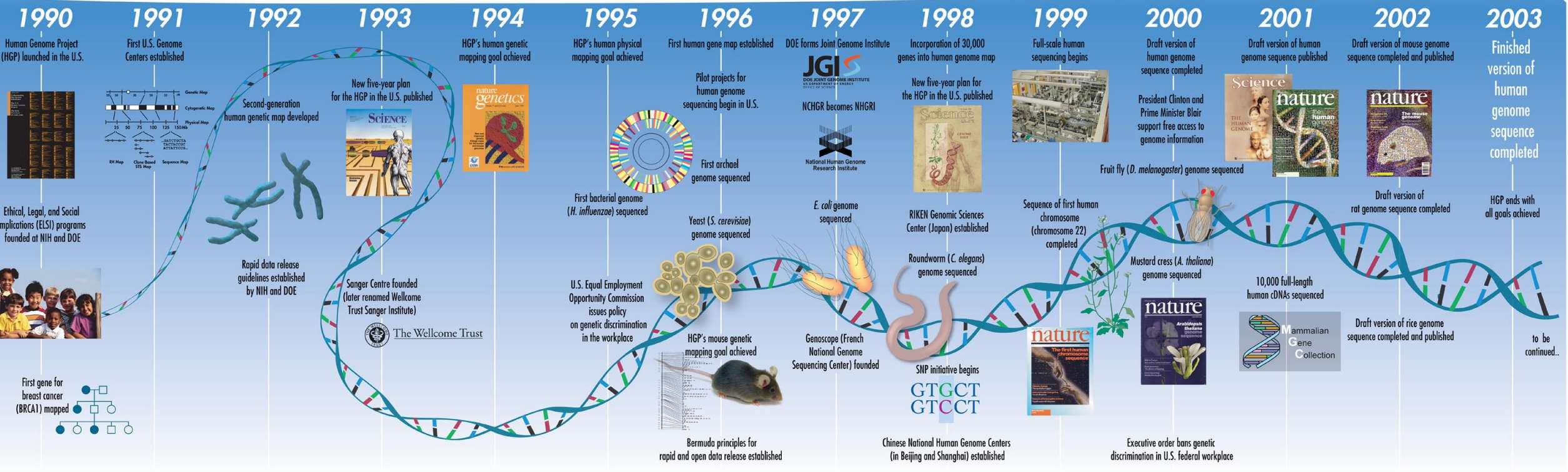
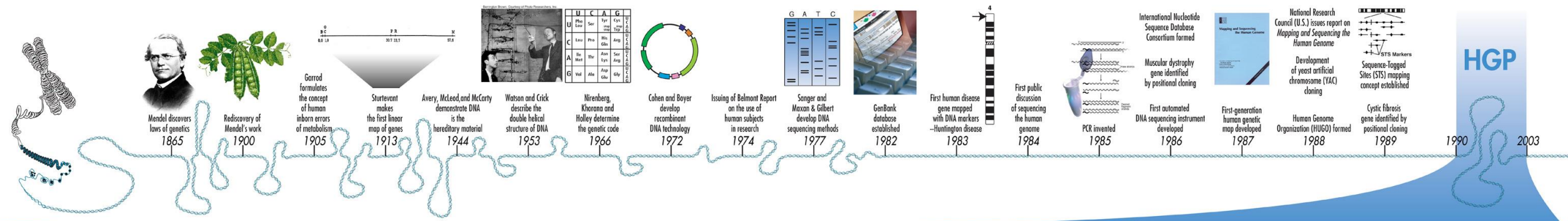
DNA PIU' LEGGERO

Effetto setaccio





# Il progetto genoma umano

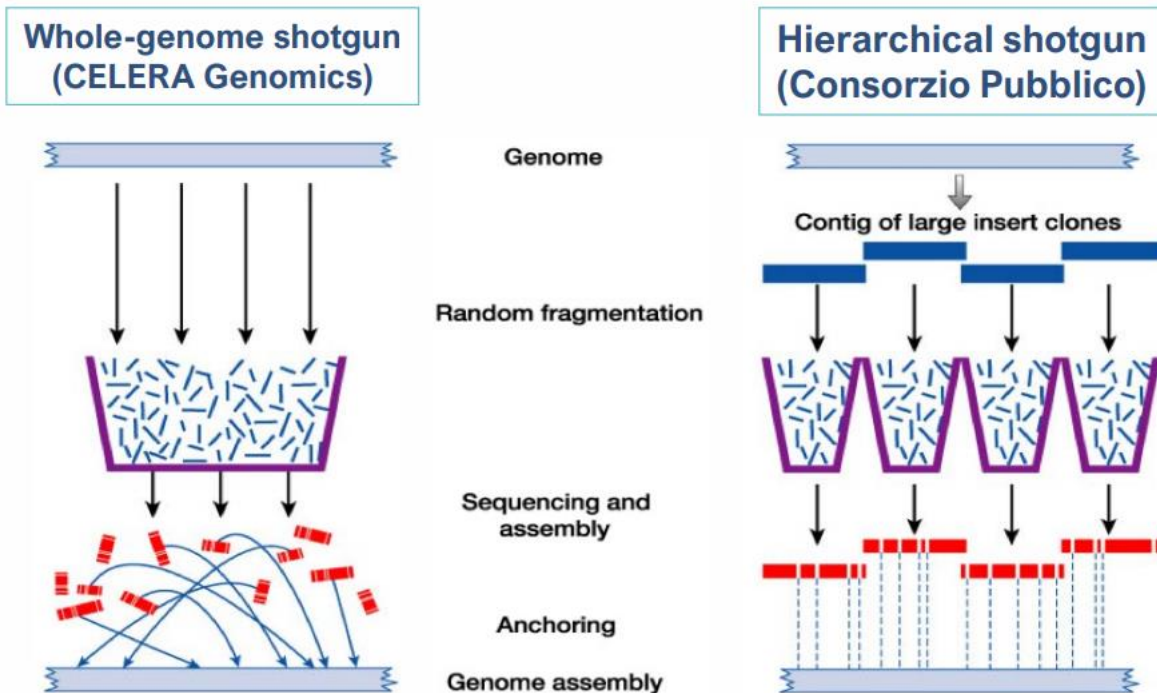


# Il progetto genoma umano

Sequenziare il genoma umano, cioè scoprire l'esatta successione dei nucleotidi presenti nel DNA rappresentativo di un essere umano (diversi volontari) in 15 anni

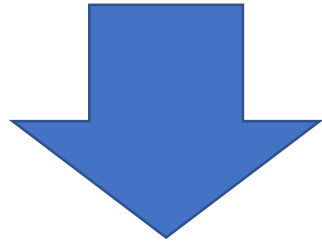
- Definire una mappa fisica e genetica del genoma umano
- Sequenziare e mappare 5 organismi modello (incluso topo)

Stabiliti nel 1988 da un US Academy of Sciences e poi ripresi dall'ente americano che lanciò il progetto (Dipartimento dell'Energia)



L'approccio dell'IHGSC prevedeva la suddivisione del genoma in diversi segmenti, minimizzandone la sovrapposizione, in base a una mappa predisposta inizialmente. I frammenti venivano poi clonati in batteri vettori capaci di replicare il DNA per poterne avere quantità sufficienti per il sequenziamento. L'allineamento delle sequenze avveniva poi per mezzo di software dedicati. La seconda tecnica, quella utilizzata dalla Celera Genomics, il whole-genome shotgun, prevedeva di frammentare il genoma in modo casuale, con parziali sovrapposizioni, in modo da permettere la ricostruzione della sequenza con analisi informatiche.

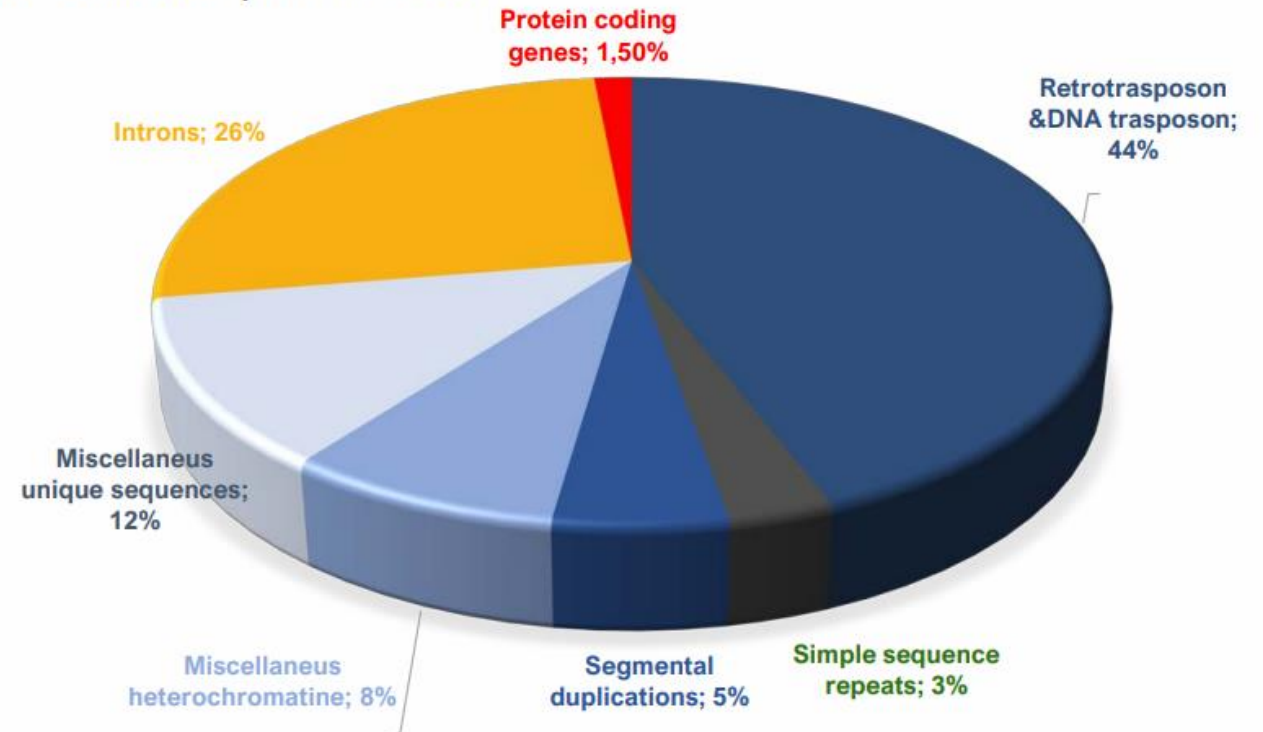
# Progetto genoma umano: i risultati



## Genomica

- Il genoma umano è lungo circa 3.2 miliardi di basi
- Possediamo tanto DNA da andare e tornare dal Sole per 41 volte!
- Possediamo «solo» 21.000 geni

La genomica si fonda sugli aspetti statici delle informazioni contenute nel DNA:  
 Quanto sono grandi i genomi  
 Quanti geni contengono  
 Come sono organizzati e distribuiti i geni nei genomi



**Sequenziamento dei genomi di altri organismi**

# Sequenziamento massivo

## GLI STRUMENTI DELLA RICERCA

1 Viene isolato un frammento di DNA a singolo filamento di cui si vuole determinare la sequenza delle basi (lo stampo).

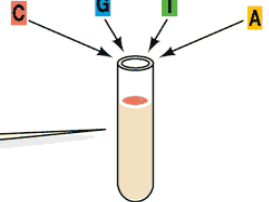
2 Ognuno dei ddNTP è legato a un diverso colorante fluorescente.

3 A un campione del DNA di cui si vuole determinare la sequenza viene aggiunto un primer, la DNA polimerasi, i dNTP e i ddNTP fluorescenti; ha inizio la sintesi.

4 Nellesempio il risultato mostra cosa si lega alla T presente nel filamento da sequenziare. Se casualmente si lega ddATP A invece di dATP, la sintesi si arresta. Viene così ottenuta una serie di frammenti di lunghezza differente, ognuno dei quali termina con un ddNTP.

5' ?????????????????????? 3'

ddCTP ddGTP ddTTP ddATP



Filamento stampo

5' ???????????????????CGCA 3'

Primer (sequenza nota)  
3' GCGT 5'

5' T??????????????????CGCA 3'

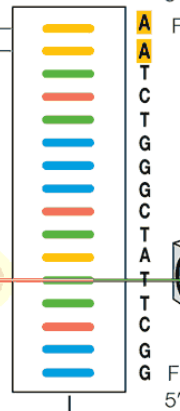
3' AATCTGGGCTATTCGGGCGT 5'

5' TT??????????????????CGCA 3'

3' AATCTGGGCTATTCGGGCGT 5'

5 I frammenti neosintetizzati di lunghezza differente vengono separati mediante elettroforesi.

Elettroforesi

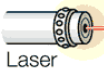


3' A  
A  
T  
C  
T  
G  
G  
C  
T  
A  
T  
T  
C  
G  
G  
5'

Frammento più lungo

Frammento più breve

6 Ogni frammento emette fluorescenza di un colore che identifica il ddNTP che ne ha determinato l'arresto della sintesi. Il colore all'estremità di ogni frammento viene rivelato per mezzo di un raggio laser.

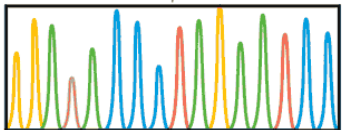


Laser

Rivelatore

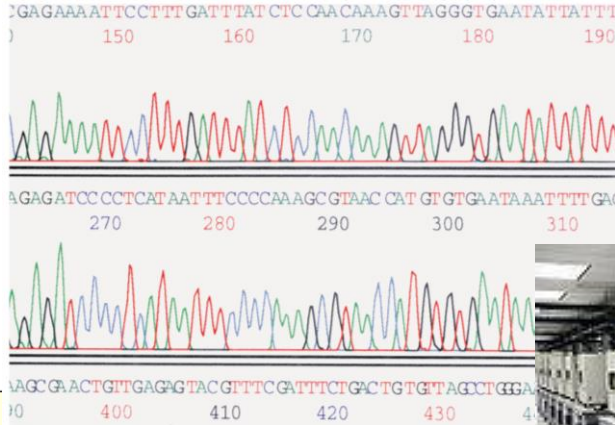
7 La sequenza del DNA può essere dedotta in base ai colori di ciascun frammento di DNA neosintetizzato...

8 ... e convertita nella sequenza complementare del filamento stampo.



3' AATCTGGGCTATTCGG 5'

5' TAGACCCGATAAGCCCGCA 3'



# Genomica comparativa

Organism	size(bp)	gene number	average gene density
<i>Homo sapiens</i>	3.1 billion	~25,000	1 gene per 100,000 bp
<i>Mus musculus</i>	2.6 billion	~25,000	1 gene per 100,000 bp
<i>D. Melanogaster</i>	137 million	13,000	1 gene per 9,000 bp
<i>A. Thaliana</i>	100 million	25,000	1 gene per 4000 bp
<i>S. Cerevisiae</i>	12.1 million	6000	1 gene per 2000 bp
<i>E.Coli</i>	4.6 million	3200	1 gene per 1400 bp

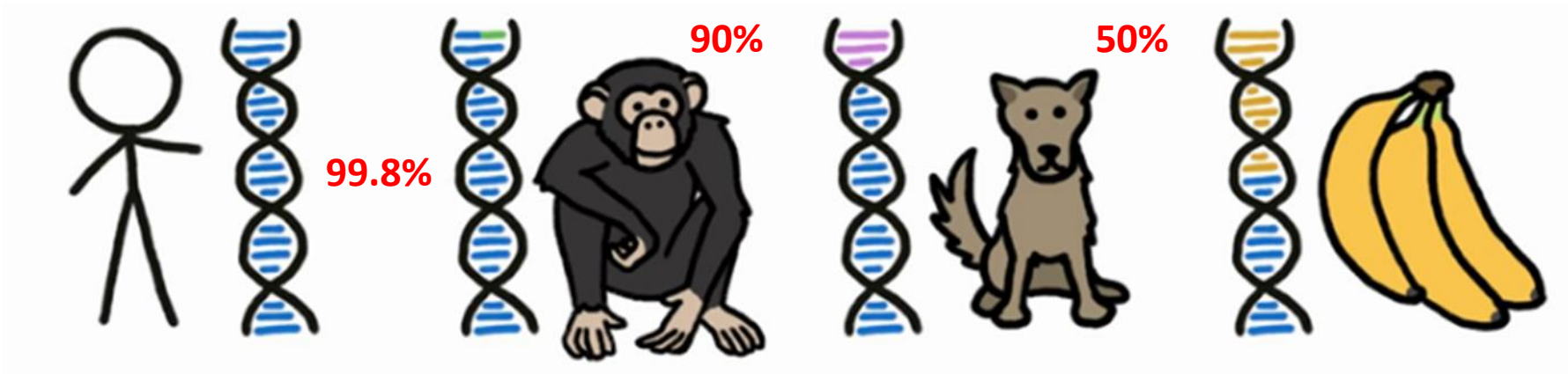
- I genomi degli organismi eucariotici sono di dimensioni superiori rispetto a quelli procariotici
- In generale, le specie pluricellulari possiedono più DNA rispetto agli eucarioti unicellulari
- I genomi eucariotici contengono un maggior numero di geni rispetto ai procarioti

**Conoscere la sequenza del genoma umano è solo il punto di partenza della ricerca genomica: ogni individuo è molto di più della semplice stringa di nucleotidi racchiusa nel suo genoma.**

Tra gli organismi pluricellulari non esiste una stretta relazione tra l'ampiezza del genoma, il numero di geni e la complessità

# Era post-genomica: Genomica funzionale

Le sequenze di DNA sono molto simili tra individui differenti!!!!



**Le differenze tra gli individui non risiedono tanto nelle singole informazioni (i geni sono conservati nel corso dell'evoluzione) ma nel modo in cui essi interagiscono**

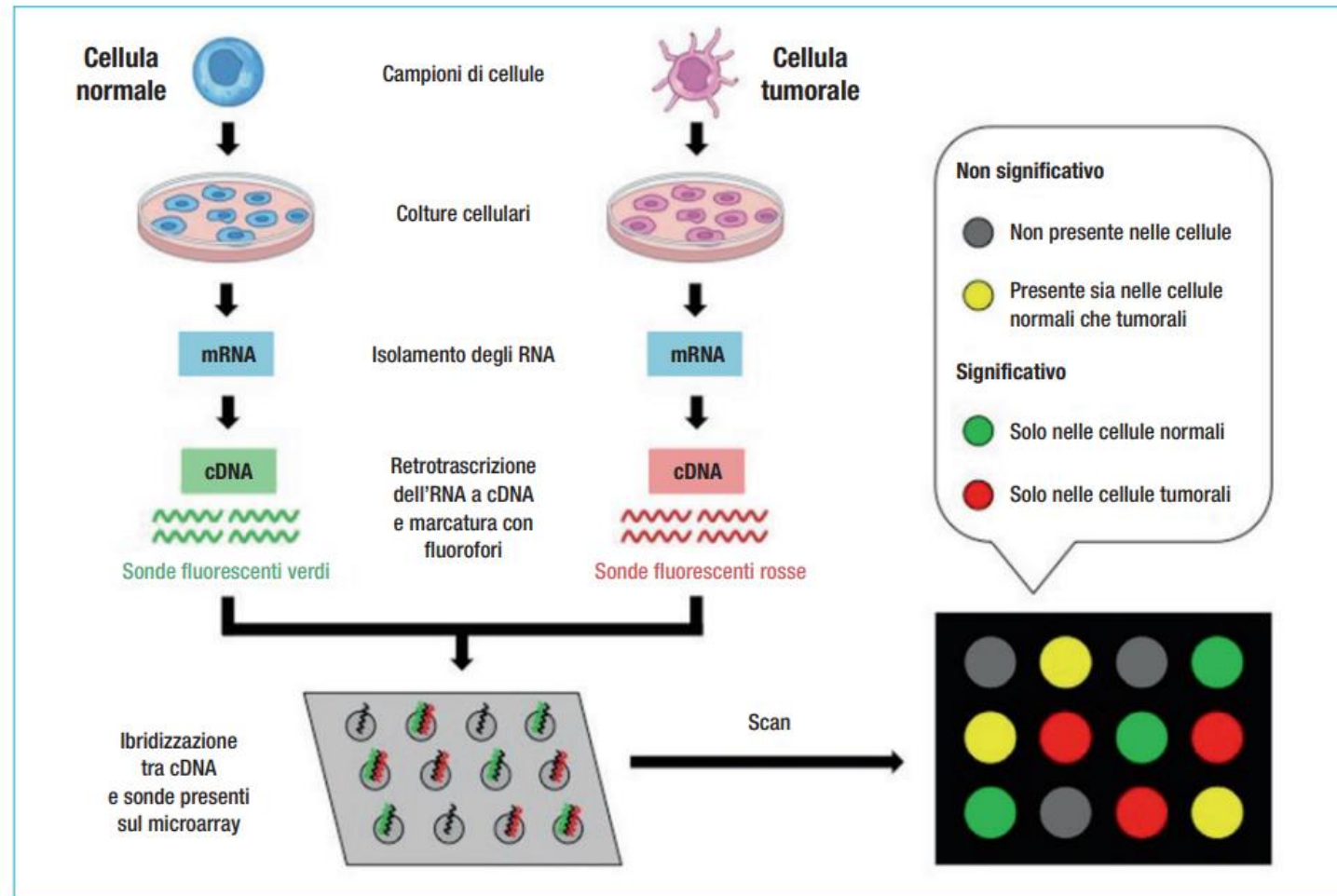


## DARE AD OGNI GENE UN SIGNIFICATO FUNZIONALE

Per comprendere la funzionalità di una cellula, il dinamismo cellulare, la plasticità cellulare, il differenziamento, l'associazione genotipo-fenotipo è necessario cambiare prospettiva, abbandonando l'approccio riduzionista, a favore di una visione olistica, in grado di descrivere tutte le interazioni che si verificano in una cellula e in un tessuto

# Era post-genomica: trascrittomica

Il trascrittoma comprende l'insieme delle molecole di **RNA messaggero** presenti in una cellula di un dato tessuto in un dato momento. Per definire un gene differenzialmente espresso si tiene conto se la sua espressione genica si discosta dalla situazione di uguale espressione nei due stati in modo significativo, confrontando, ad esempio, con un valore soglia per definire se sovrao sotto-espressi rispetto a tale valore di espressione.

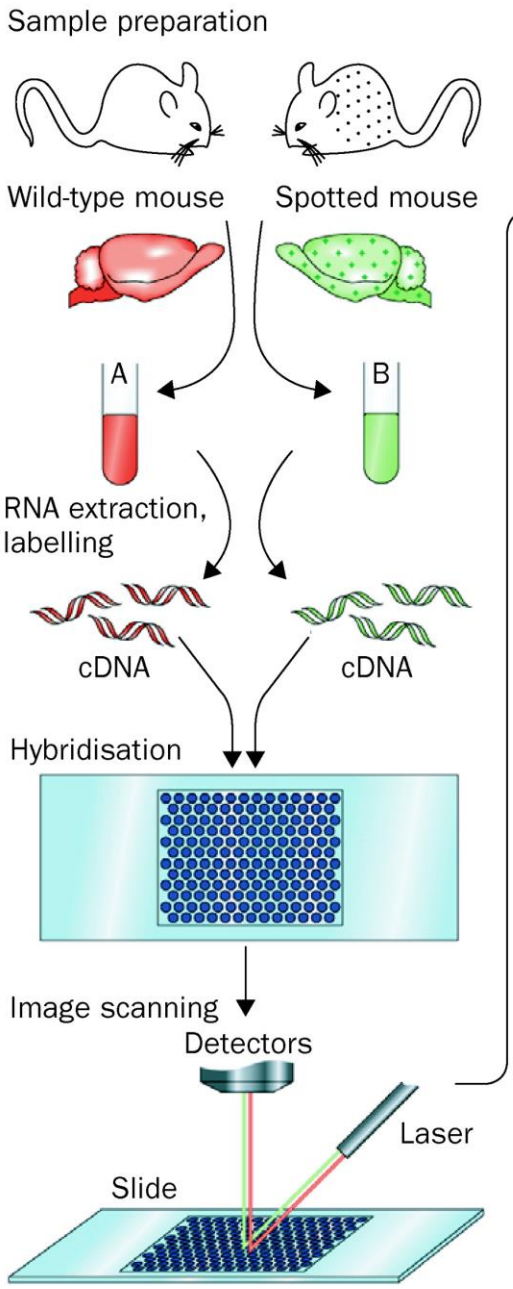


L'analisi dei trascrittomi viene effettuata attraverso l'utilizzo di due differenti tecnologie:

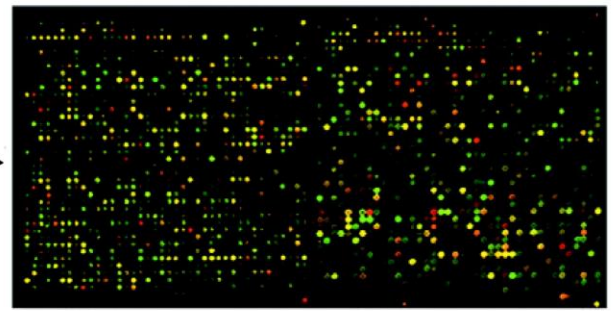
- **Microarray**
- **NGS (Next-Generation Sequencing).**

1995 vennero sviluppati i primi microarray basati su spotting di molecole di cDNA,  
2002 si passò a utilizzare i cosiddetti high density oligo microarrays;  
2008 la tecnica di RNA-Seq permette il sequenziamento degli RNA messaggeri attraverso tecniche di tipo NGS

# La tecnologia dei microarray



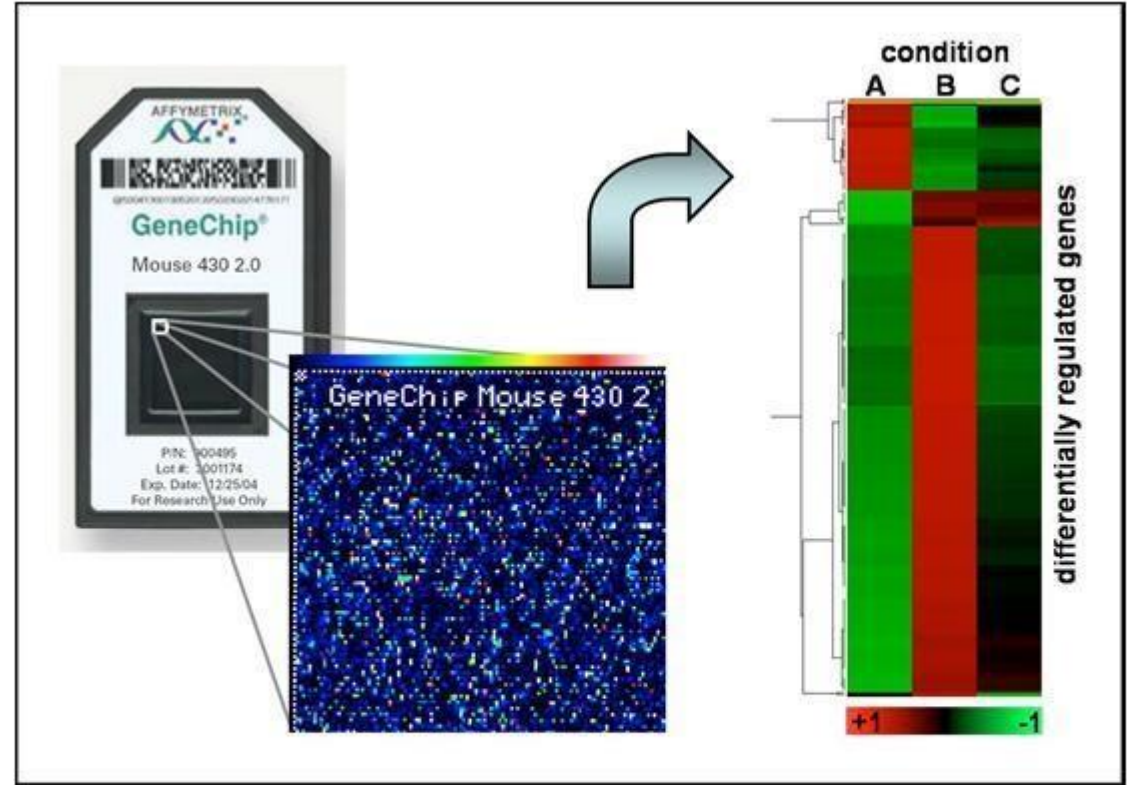
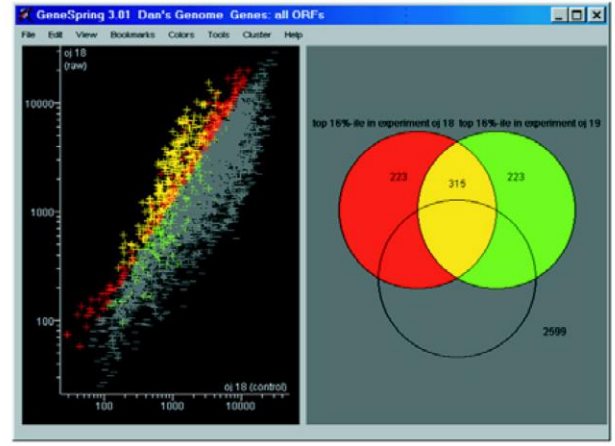
False colour image representing greyscale values



Data analysis

Signif. Data	File	Meta	Meta	Meta	Column	Gene	Reg	Sigral Mean	Background	Sigral	Background	Sigral	Background	Sigral	Sigral	Sigral
A	1	1	1	1	2	2	1	117.2261	87.2311	396	158	89	261	2032	7090	27.211
A	1	1	1	1	2	2	1	1302.482	124.753	254	89	66	10092	7090	267.202	
A	1	1	1	1	2	2	1	2119.658	113.028	1144	112	228	20675	22173	2142.52	
A	1	1	1	1	2	2	1	1506.962	146.721	192	192	59	25962	21724	891.976	
A	1	1	1	1	2	2	1	2871.864	1762.479	2740	1748	107	813	209718	182142	288.104
A	1	1	1	1	2	2	1	2746.012	180.012	1744	1862	140	718	474968	146210	101.174
A	1	1	1	1	2	2	1	2034.128	2222.655	2077	2226	74	725	271284	1674188	402.175
A	1	1	1	1	2	2	1	4487.177	1676.764	4488	2647	114	857	758664	1308701	715.887
A	1	1	1	1	2	2	1	4845.366	2832.78	4282	2742	133	749	844504	212775	118.412
A	1	1	1	1	2	2	1	177.084	87.813	396	158	89	261	2032	7090	27.211
A	1	1	1	1	2	2	1	1302.482	124.753	254	89	66	10092	7090	267.202	
A	1	1	1	1	2	2	1	2119.658	113.028	1144	112	228	20675	22173	2142.52	
A	1	1	1	1	2	2	1	1506.962	146.721	192	192	59	25962	21724	891.976	
A	1	1	1	1	2	2	1	2871.864	1762.479	2740	1748	107	813	209718	182142	288.104
A	1	1	1	1	2	2	1	2746.012	180.012	1744	1862	140	718	474968	146210	101.174
A	1	1	1	1	2	2	1	2034.128	2222.655	2077	2226	74	725	271284	1674188	402.175
A	1	1	1	1	2	2	1	4487.177	1676.764	4488	2647	114	857	758664	1308701	715.887
A	1	1	1	1	2	2	1	4845.366	2832.78	4282	2742	133	749	844504	212775	118.412

Data mining/database



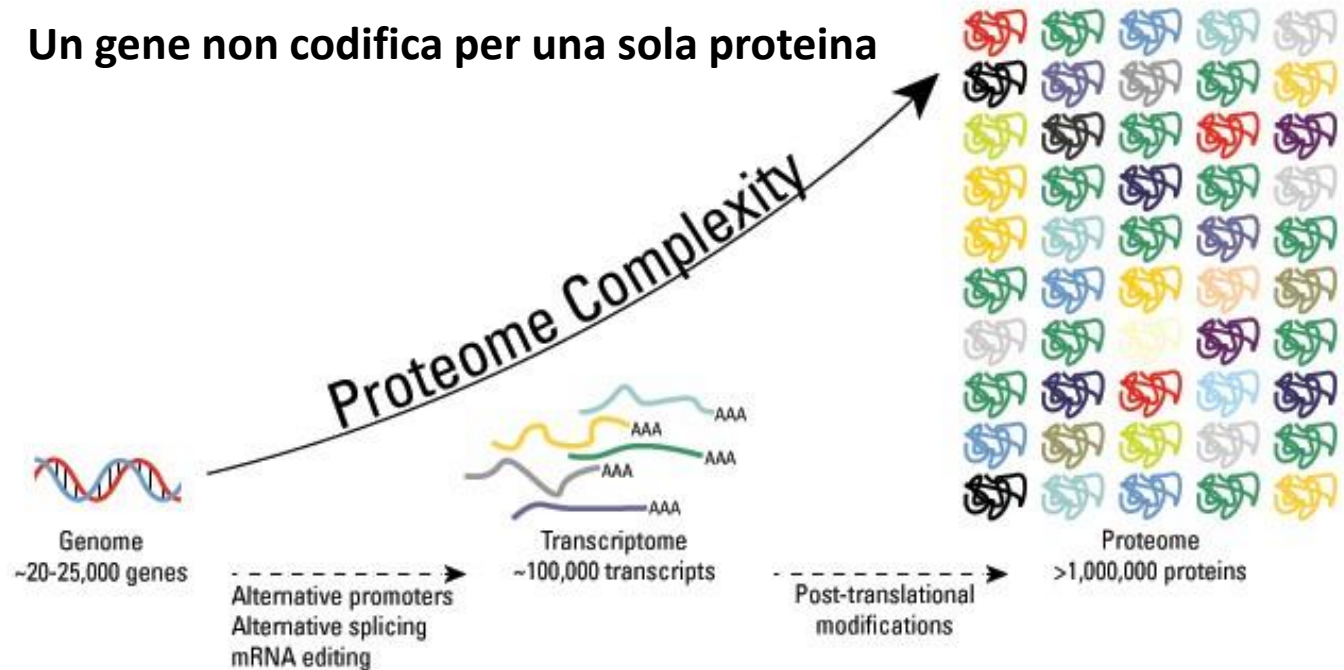
Un microarray è un insieme di microscopiche sonde di DNA legate a una superficie solida (vetro, plastica o chip di silicio) formando un array (matrice). Tali array permettono di esaminare simultaneamente la presenza di moltissimi geni all'interno di un campione di DNA. I limiti di questa tecnica sono numerosi: prima di tutto è necessario conoscere la sequenza dei geni di cui valutare l'espressione, al fine di disegnare le probe da caricare sul microarray; l'ibridazione è inoltre influenzata da "rumore di fondo" dato da appaiamenti incompleti, imprecisi, che riducono la specificità della tecnica;



# Lo studio del proteoma è molto più complesso dello studio del genoma

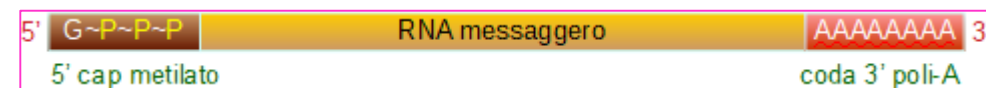
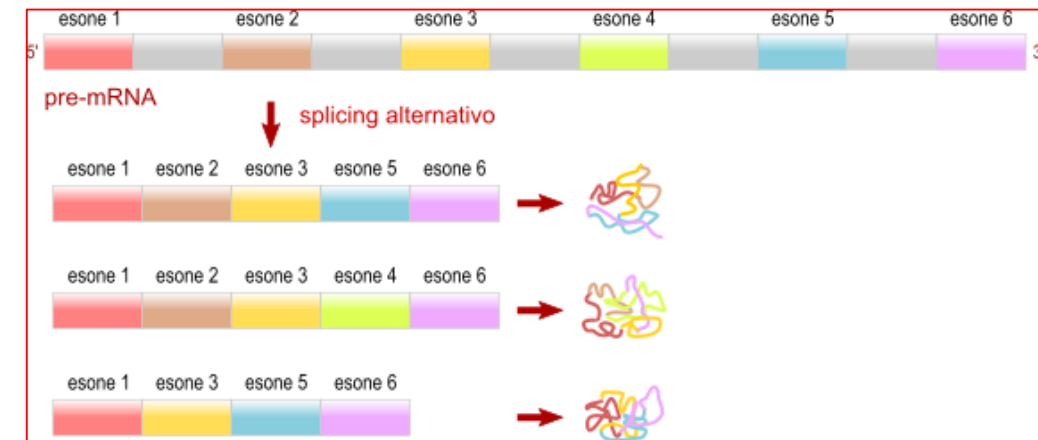
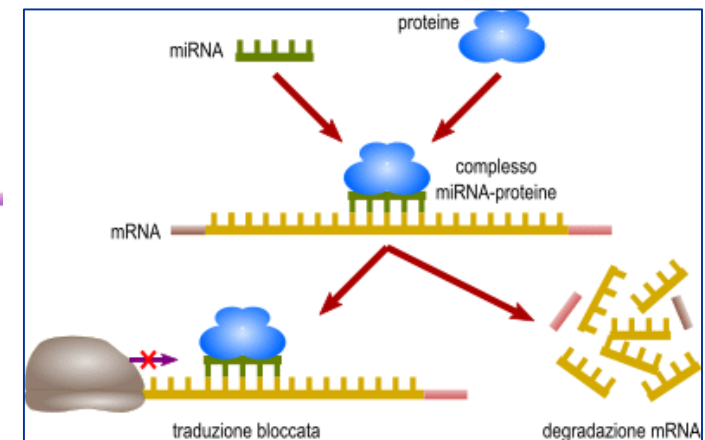
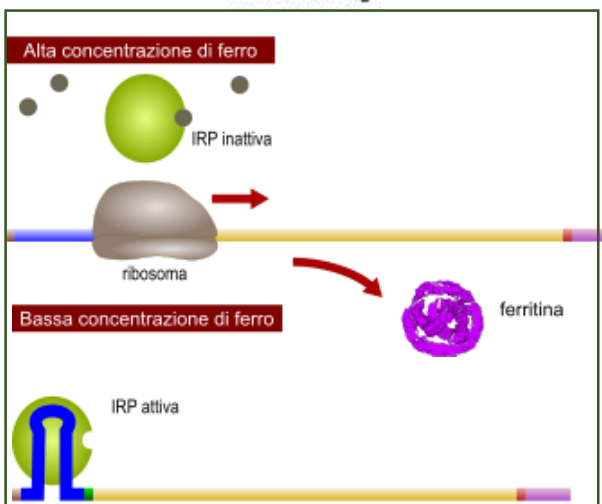
Scienze omiche: Grandi apparecchiature per la Proteomica

Un gene non codifica per una sola proteina



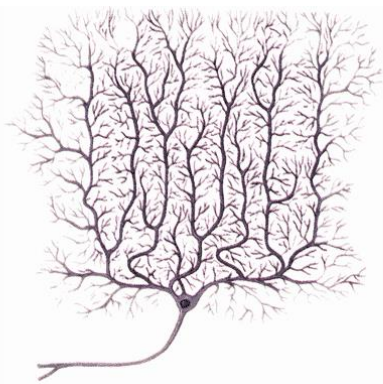
Non vi è una correlazione lineare tra quantità di mRNA e proteina

- Splicing e splicing alternativo
- Editing dell'RNA
- Regolazione genica post-trascrizionale da parte dei microRNA
- Controllo di qualità ed esportazione nucleare dei mRNA
- Il turnover dell'RNA citoplasmatico
- Traducibilità e induzione/repressione della traduzione

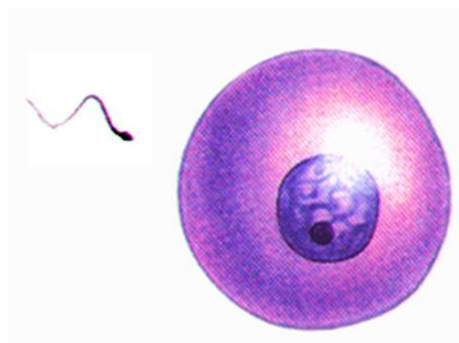
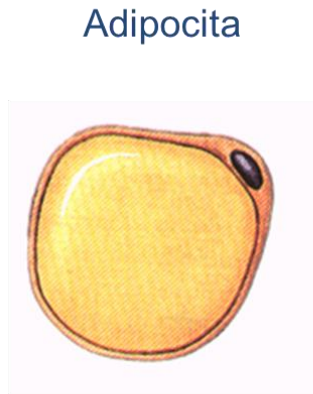


# Lo studio del proteoma è molto più complesso dello studio del genoma

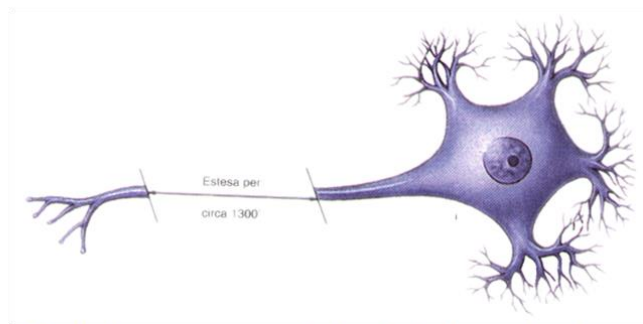
Mentre il genoma è costante per una data cellula ed identico per tutte le cellule di un organismo (e non cambia molto all'interno della specie) il proteoma è dinamico nel tempo, si modifica in risposta a fattori esterni e differisce in maniera sostanziale tra i diversi tipi cellulari



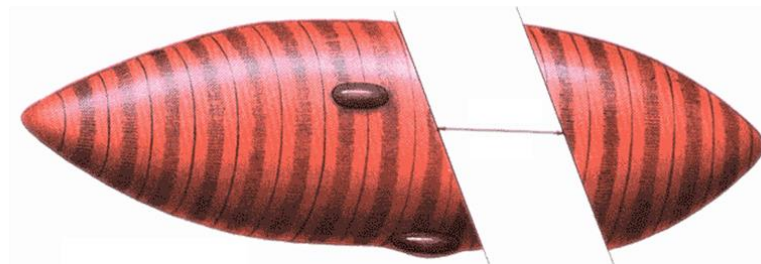
Cellula nervosa del cervelletto



Cellula uovo e spermatozoo

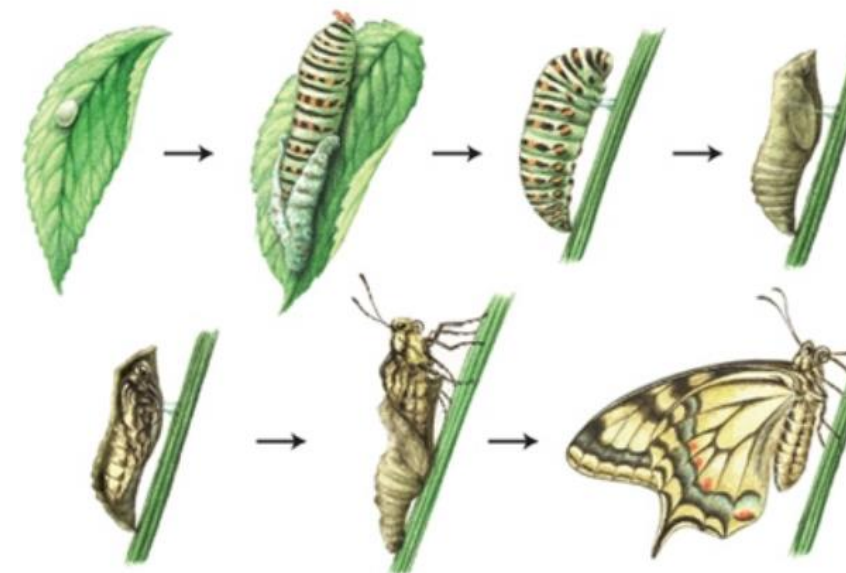


Cellula nervosa del midollo spinale



Cellula muscolare scheletrica

## Correlazione genotipo-fenotipo

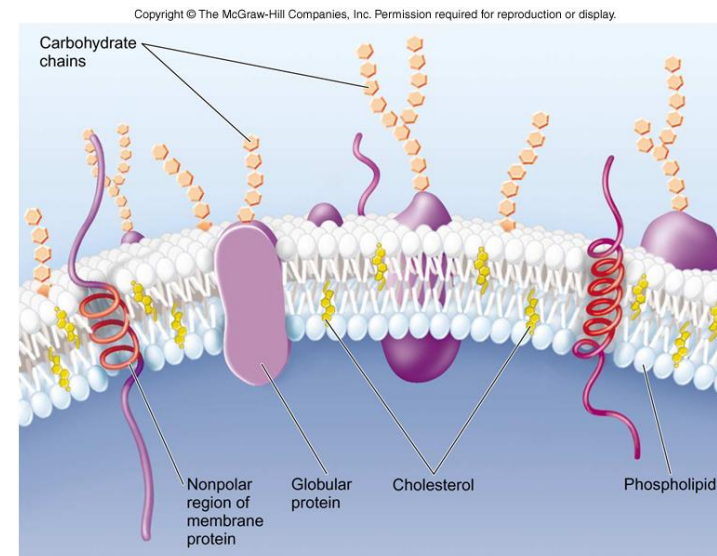
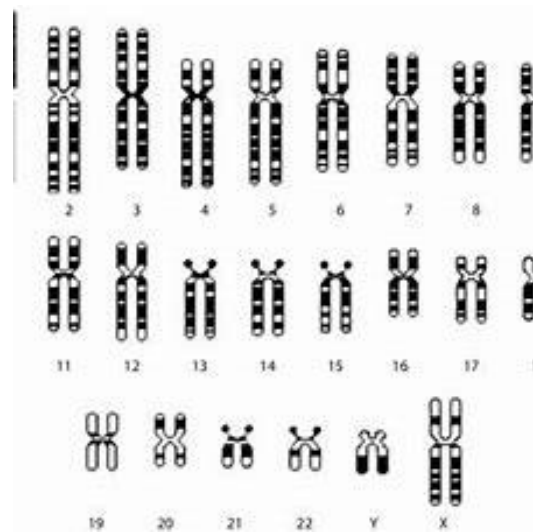
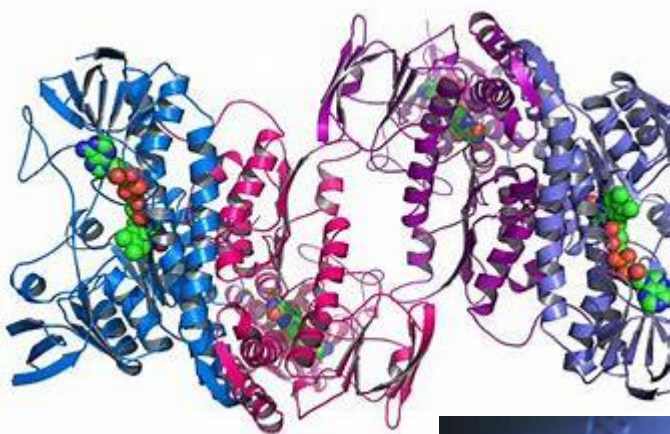


Stesso genoma, differente proteoma, differente fenotipo

# Lo studio del proteoma è molto più complesso dello studio del genoma

Il materiale genetico è formato dalla ripetizione di 4 basi azotate, la complessità strutturale delle proteine è notevolmente più elevata (20 tipi di amminoacidi e coniugati con macromolecole differenti come lipidi, carboidrati, nucleotidi)

Il DNA è presente in quantità uniforme in una cellula, il livello di espressione quantitativo delle proteine che esso codifica può variare in un intervallo di concentrazione da 1 a  $1 \times 10^6$  /  $1 \times 10^9$



# Quali sono le domande alle quali la Proteomica può dare una risposta

**1. Quali proteine sono differenzialmente espresse oppure modificate (sia in termini qualitativi che quantitativi) confrontando due (o più) fenotipi cellulari**

Un esempio di diversi fenotipi cellulari potrebbe essere:

- Cellule di origine diversa;
- Cellule dello stesso tipo sottoposte a stimoli diversi;
- Cellule dello stesso tipo soggette o meno a condizioni patologiche.

**2. Quali proteine sono differenzialmete soggette ad una particolare PTM confrontando fenotipi cellulari diversi**

**Fosfo**proteoma  
**Metil**proteoma  
**Acetil**proteoma  
**Ubiquitin**proteoma  
**Glicosil**proteoma ...

**3. Quali proteine costituiscono/sono presenti in un determinato complesso macromolecolare/organello**

Proteoma nucleare  
Proteoma di membrana  
Proteoma mitocondriale...

**4. Quali proteine interagiscono con una determinata proteina (X)  
Determinazione del network molecolare di una proteina (X)**

# PROTEOMICA

Un settore emergente della ricerca delle scienze della vita che usa tecnologie High Throughput (HT) per la visualizzazione, identificazione e/o caratterizzare tutte le proteine in una determinata cellula, tessuto o organismo in un determinato momento (i.e. il proteoma). La proteomica si prefigge inoltre la determinazione della loro localizzazione, le modifiche, interazioni, le attività, e, in ultima analisi, la loro funzione.

## Finalità della proteomica

### 1. Proteomica strutturale

1. Sito catalitico di una proteina
2. Protein-protein interaction
3. Stabilire i pathways metabolici

### 2. Proteomica funzionale

1. Modifiche post traduzionali (PTM)
2. Localizzazione subcellulare
3. Interazioni (Interactome)
4. Mutazioni (Diseases)

COMPRENDERE COME UNA CELLULA REGOLA LE ATTIVITA' CELLULARI

## Metodologie proteomiche

Tecniche di Biologia molecolare

Cromatografia

Bioinformatica

Elettroforesi

Robotica

Cristallografia ai raggi X

Spettrometria di massa

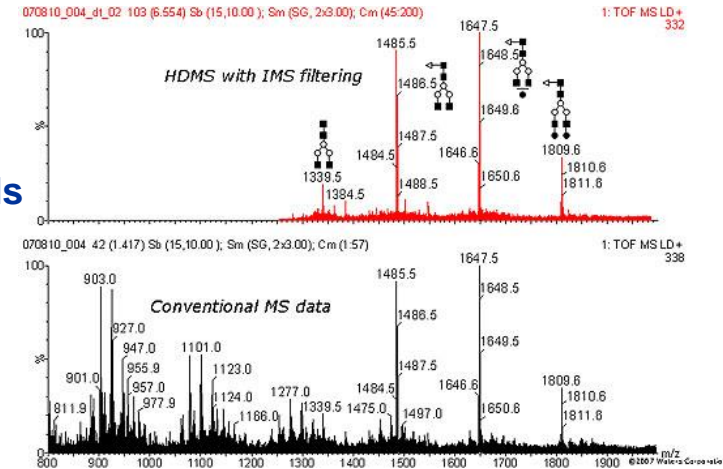
Spettroscopia NMR

# Scelta della piattaforma proteomica

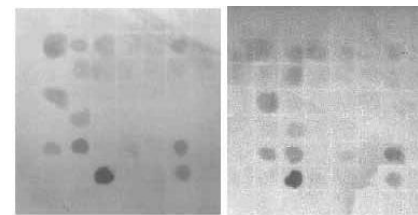
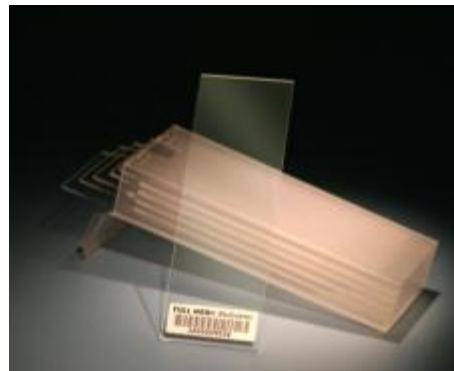


Two-dimensional gel electrophoresis (2-DE) of intact proteins: 2D-IPG; 2D-DIGE

MS-based (“gel-free”) methods  
“Shotgun” LC-ESI-MS/MS of total tryptic digest of proteins  
Quantitation by stable isotope labelling  
(e.g. SILAC, ICAT, iTRAQ)



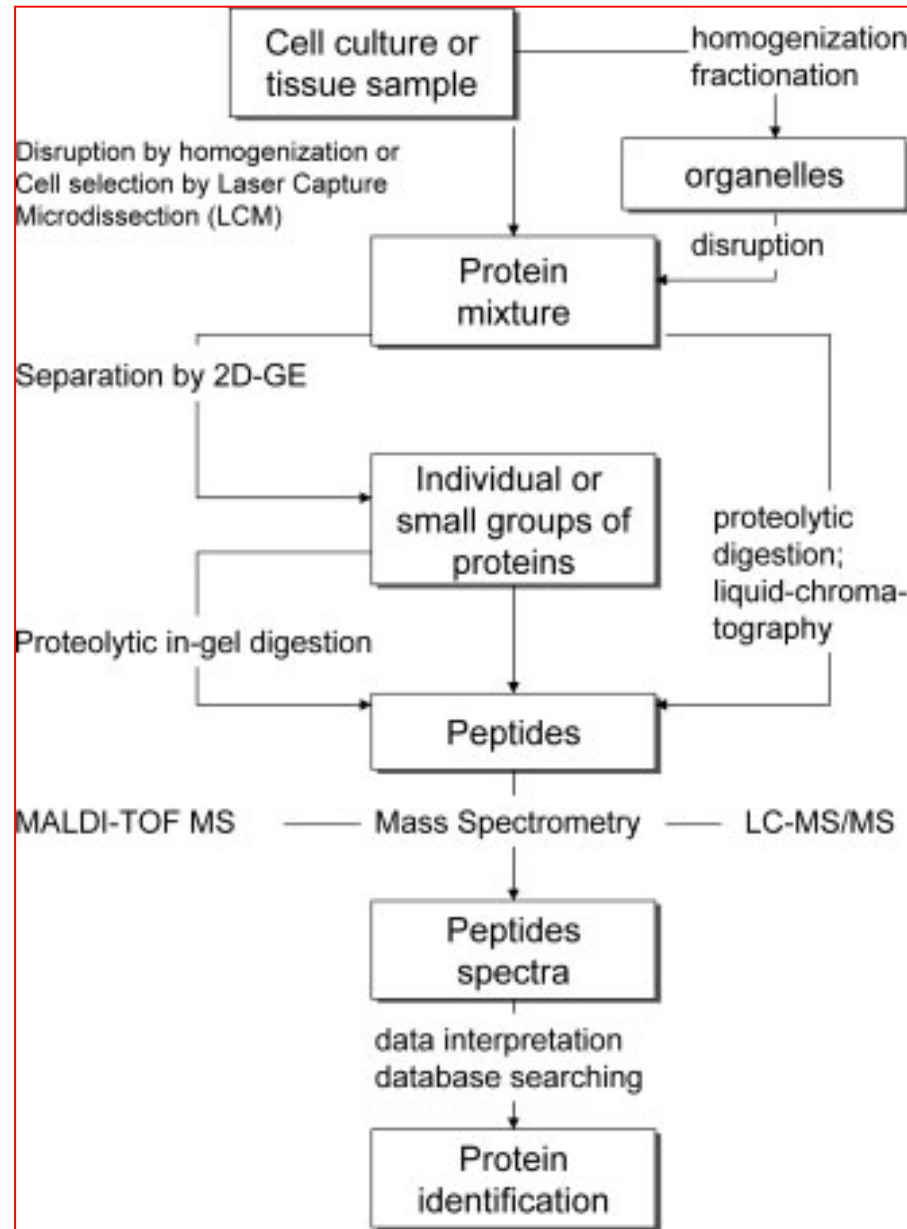
Protein chips  
Intact proteins  
e.g. SELDI-MS, protein,  
tissue  
and antibody arrays)



# DAL CAMPIONE AL PROTEOMA: UN PROTOCOLLO INTERNAZIONALE

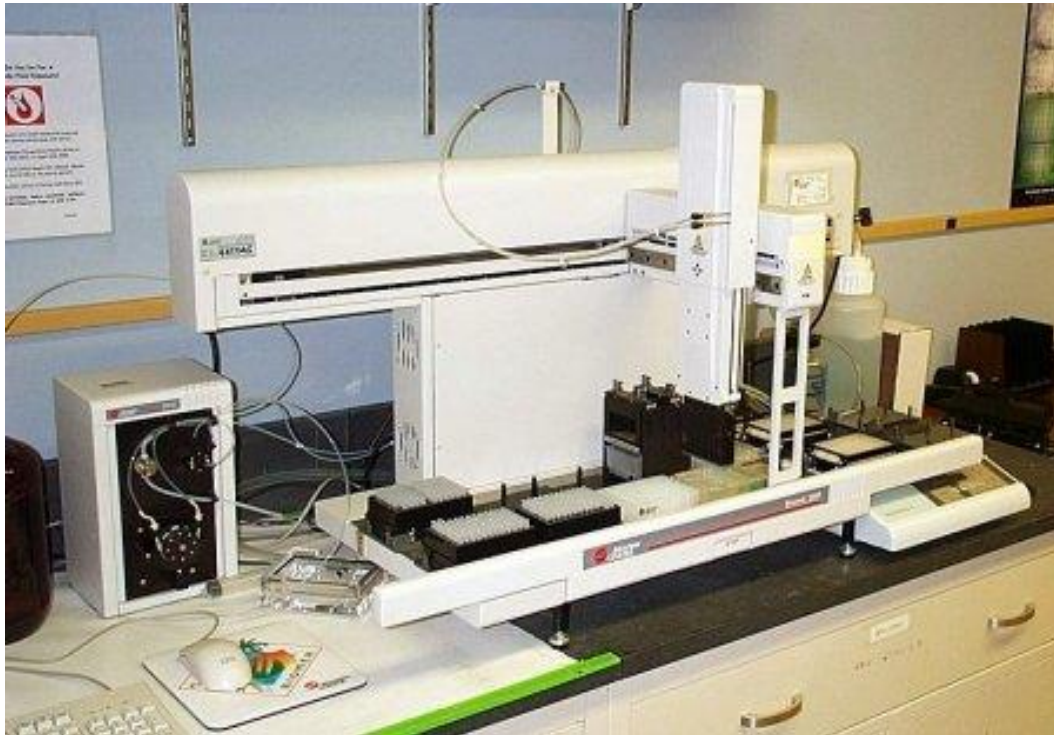
## WORKFLOW

1. Sample preparation
2. Protein separation
3. Protein identification



# TECNOLOGIE PROTEOMICHE DI SECONDA GENERAZIONE

1. Large scale
2. High-throughput



**Workstations for sample preparation, spot picking, in-gel digestion of proteins, identification by PMF, database searching and data mining.**



# Proteomics Workflow

2D Separation

Sample Prep



Staining

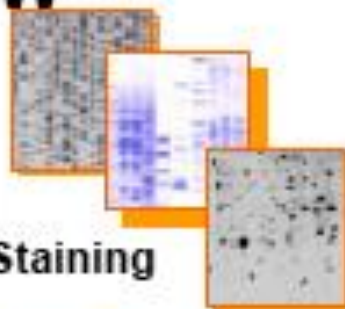


Image acquisition

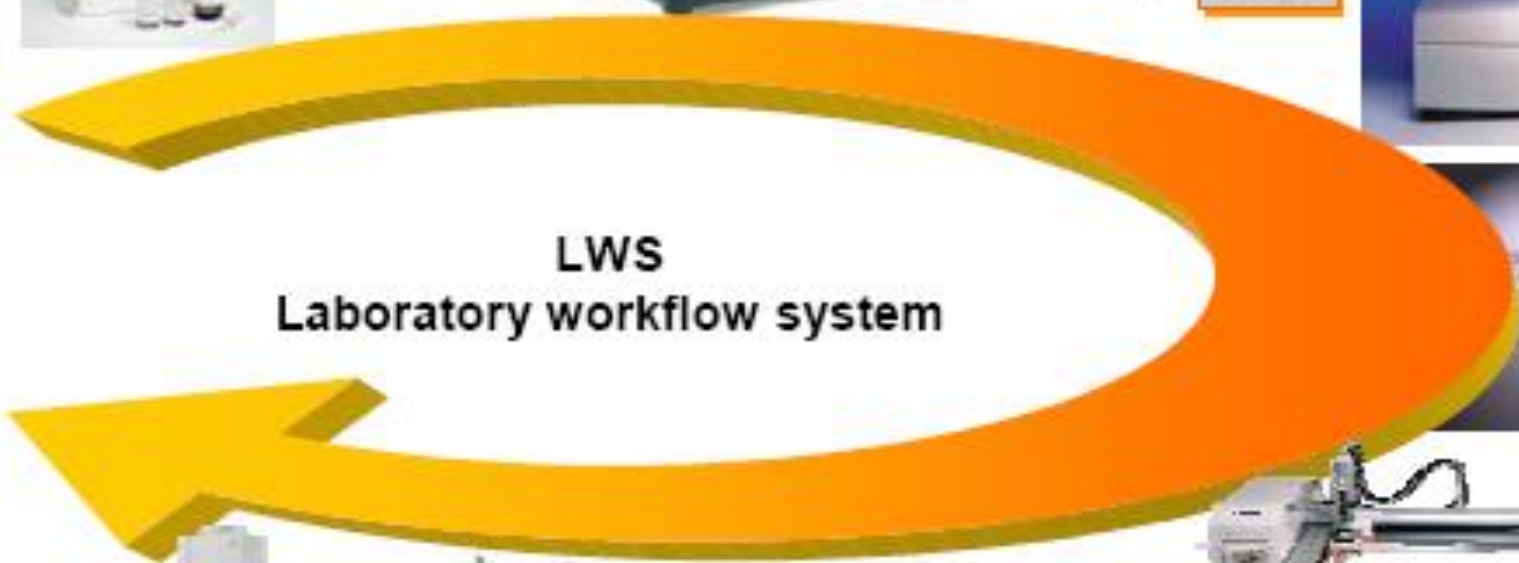


Image analysis



MALDI-ToF



MALDI spotting



Spot digestion



Automated spot picking



## Aree della proteomica che necessitano di miglioramenti

- \* **Sample preparation and solubilization**

- **Insoluble sample: membrane and nuclear proteins**

- **Proteins from highly resistant tissues like hair and skin**

- \* **Low abundance proteins**

- \* **Prefractionation**

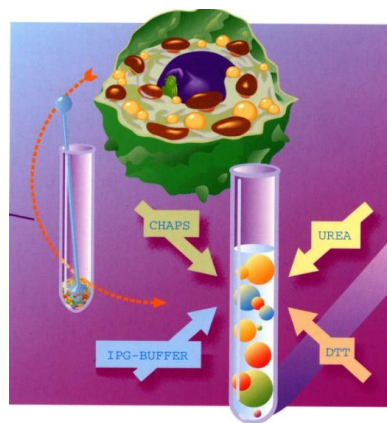
- \* **High sample loads**

- \* **Basic proteins**

- \* **Quantitation**

# Le fasi dello studio proteomico gel- based

- Preparazione del campione
- Elettroforesi bidimensionale 2D-DIGE
  - Prima dimensione –IPG–*
  - Seconda dimensione SDS-PAGE*
- Rivelazione delle proteine sulla mappa bidimensionale
- Analisi computerizzata delle mappe proteomiche
- Identificazione degli *spots* proteici
  - Gel matching*
  - Western Blot*
  - Microsequenziamento N-terminale*
  - Spettrometria di massa (MALDI-TOF e Ms/Ms)*
- Analisi statistica dei dati



## Distruzione delle cellule

### METODI MECCANICI

- omogeneizzatori
- Mulini a sfere
- French press
- sonicatori

### 1. OMOGENIZZAZIONE



Il tessuto viene messo in una provetta di vetro (in cui si aggiunge il buffer di estrazione) all'interno del quale agisce un pestello di vetro o di metallo (Potter) che può essere azionato a mano o mediante un motore elettrico



### METODI NON MECCANICI

- Shock osmotico
- Congelamento e scongelamento
- Enzimi litici

### METODI NON MECCANICI DI FRANTUMAZIONE

**DIGESTIONE ENZIMATICA** (es. lisozima per la parete batterica) proteolisi limitata  
**DETERGENTI O SOLVENTI ORGANICI** (SDS, Triton X-100, acetone, toluene). Attenzione: potrebbero denaturare la proteina di interesse. Necessari per proteine di membrana  
**CONGELAMENTO/SCONGELAMENTO**. Attenzione: alcune proteine potrebbero denaturarsi o degradarsi  
**SHOCK OSMOTICO** (soluzioni ipotoniche): il metodo più delicato. Va bene se la proteina è citosolica



In una soluzione, **pressione osmotica (π)**, **osmolarità (OST)** e **osmolalità** sono proprietà correlate. Il valore di osmolarità, che corrisponde approssimativamente all'osmolarità, è la somma delle concentrazioni di tutti i soluti ad una particolare temperatura. Osmolarità e pressione osmotica si possono misurare, ma non calcolare

Possono influenzare la stabilità di una proteina:

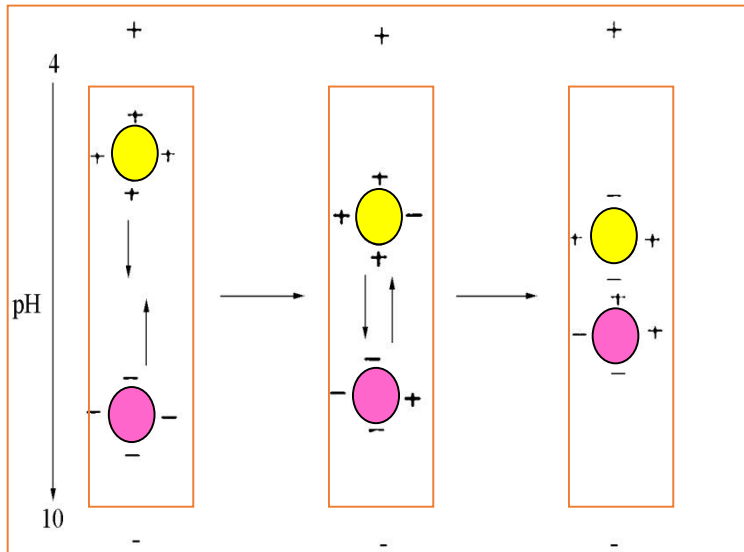
- pH
- temperatura
- concentrazione salina
- sensibilità all'ossigeno
- forze meccaniche
- condizioni di conservazione

I tamponi di solubilizzazione comunemente utilizzati contengono alcuni componenti basilari per la funzione stessa del tampone che sono:

- a) agenti denaturanti
- b) detergenti
- c) agenti riducenti
- d) anfotili carrier.

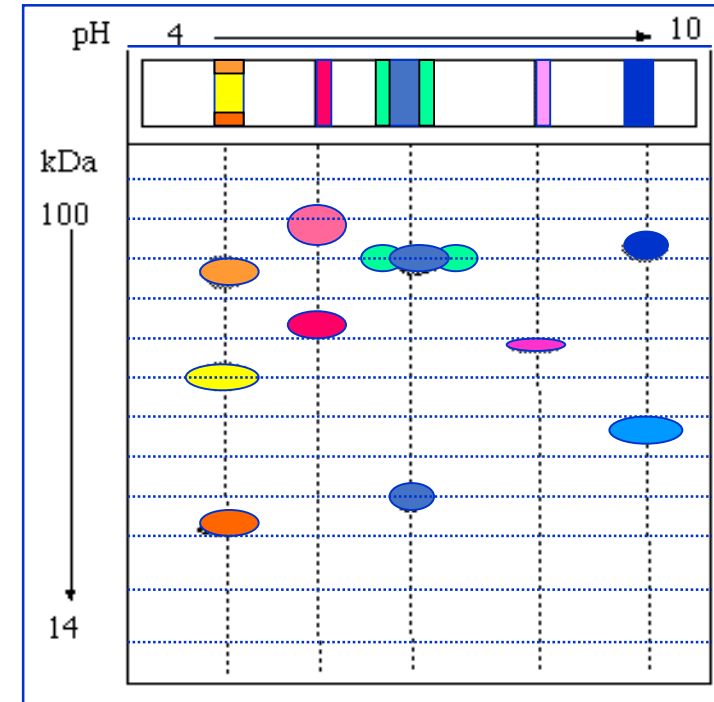
# L'elettroforesi bidimensionale

## I dimensione



Schema di separazione di proteine su un gradiente di pH (1° dimensione). Il metodo sfrutta il comportamento anfotero delle proteine, che a pH acido si caricano positivamente, a pH basico negativamente; al pH del loro punto isoelettrico (pI) sono elettricamente neutre.

## II dimensione

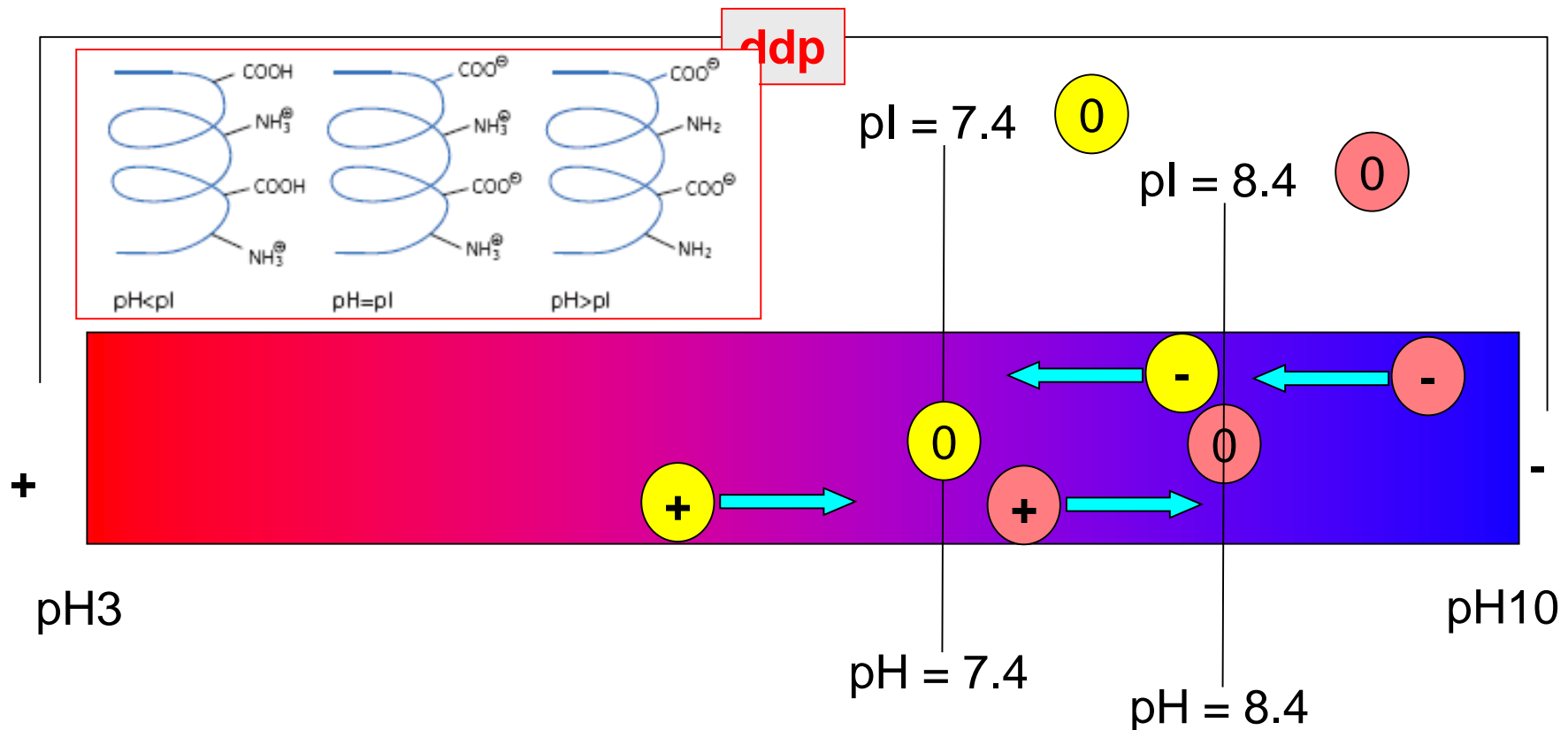


Schema di separazione di proteine in seconda dimensione, in base al loro Peso Molecolare (in presenza di SDS).

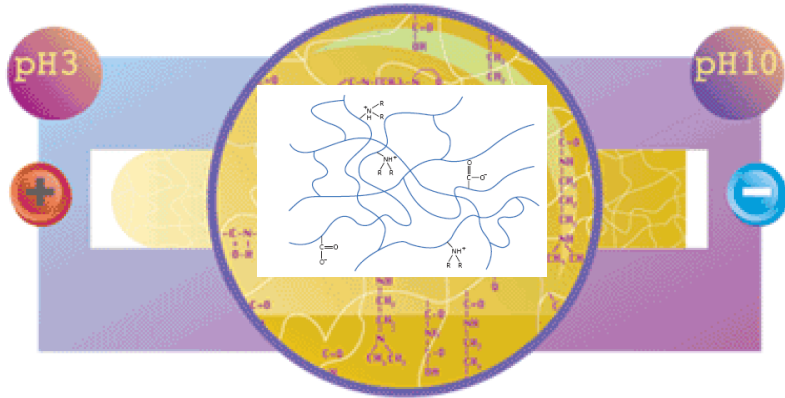
# I dimensione: IPG

## (Immobilized pH Gradient Electrophoresis)

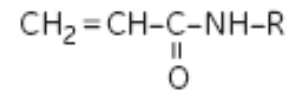
Una proteina dispersa in un gradiente di pH, si troverà ad avere carica netta positiva, negativa oppure nulla (se si trova già ad un pH pari al suo pI. Sottoposta all'azione di un campo elettrico **opportunamente orientato** essa si muoverà, a seconda della carica che reca verso l'elettrodo si segno opposto, fino a raggiungere il pH pari al suo pI. In questo punto essa assume carica netta nulla e non è più sottoposta all'azione del campo elettrico.



# IPG: Immobilized pH Gradient Electrophoresis

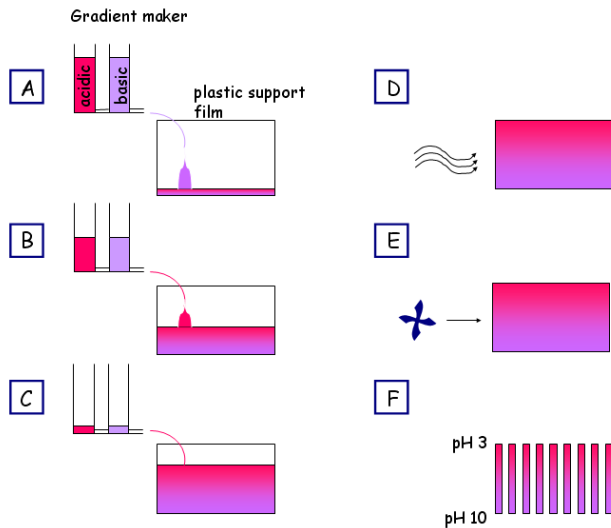


Le immobiline (derivati carichi della poliacrilamide) non sono anfotere, ma bifunzionali: hanno cioè ad una estremità lo ione tamponante e all'estremità opposta il doppio legame acrilico, che permette loro di ancorarsi alla matrice di poliacrilamide.



*R = weakly acidic or basic buffering group*

## Production of Immobilized pH Gradient (IPG) strip



Esistono in commercio gel di immobiline che coprono svariati range di pH. I gel sono venduti disidratati e congelati a  $-20^{\circ}\text{C}$

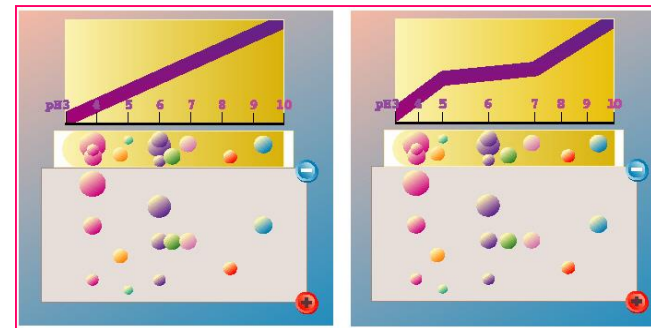


### Range di pH:

3-10 L. e NL., 4-7, 6-11,  
3.5-4.5, 4-5, 4.5-5.5, 5-6

### Dimensioni:

7, 11, 18, 24 cm



Gradiente 3-10 Lineare

Gradiente 3-10 Non-Lineare

# Ettan IPGphor II Isoelectric Focusing System

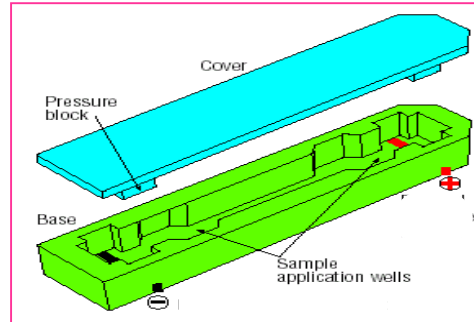


Fig 26. Applying rehydration solution into the Strip Holder.

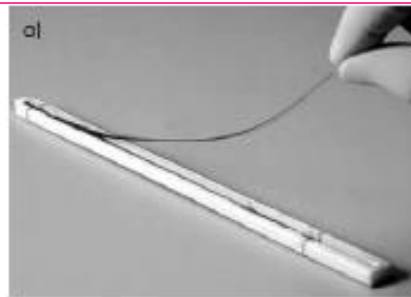


Fig 27. a, b. Positioning the Immobiline DryStrip gel.

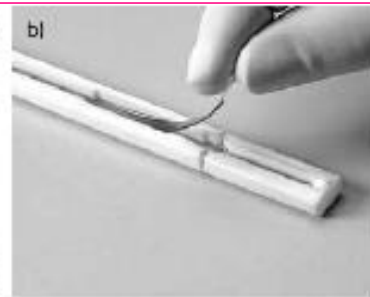


TABLE 16. IMMOBILINE DRYSTRIP IEF GUIDELINES FOR IPGPHOR ISOELECTRIC FOCUSING SYSTEM

	Step	Voltage	(h:min)	(Vh)	type
7 cm		rehydration	12:00 <sup>1</sup>		
	1	500	0:30	250	Step-n-hold
	2	1000	0:30	500	Step-n-hold
	3	8000 <sup>2</sup>	3:45	30000	Step-n-hold
11 cm		rehydration	12:00 <sup>1</sup>		
	1	500	1:00	500	Step-n-hold
	2	1000	1:00	1000	Step-n-hold
	3	8000 <sup>2</sup>	7:30	60000	Step-n-hold
13 cm		rehydration	12:00 <sup>1</sup>		
	1	500	1:00	500	Step-n-hold
	2	1000	1:00	1000	Step-n-hold
	3	8000 <sup>2</sup>	9:30	75000	Step-n-hold
18 cm		rehydration	12:00 <sup>1</sup>		
	1	500	1:00	500	Step-n-hold
	2	1000	1:00	1000	Step-n-hold
	3	8000 <sup>2</sup>	12:30	100000	Step-n-hold
	4	8000 <sup>2</sup>	2:30	20000	Step-n-hold

<sup>1</sup> The total rehydration time can be adjusted somewhat for convenience, but must be greater than 10 hours.

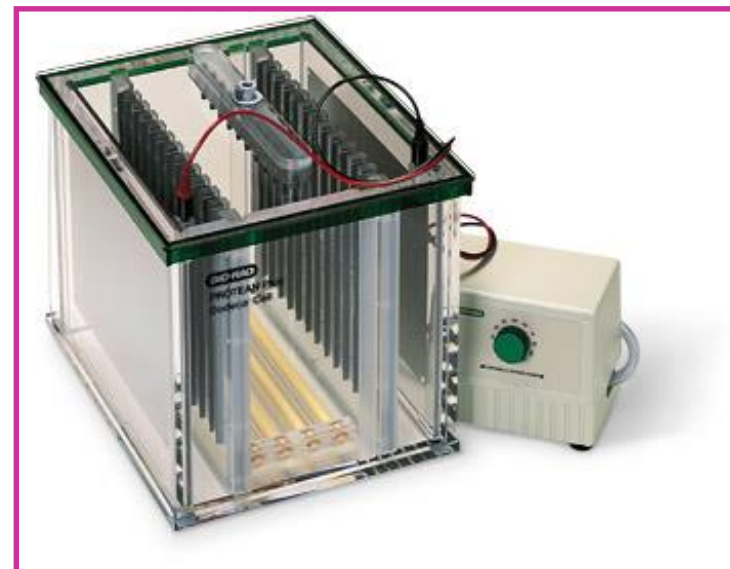
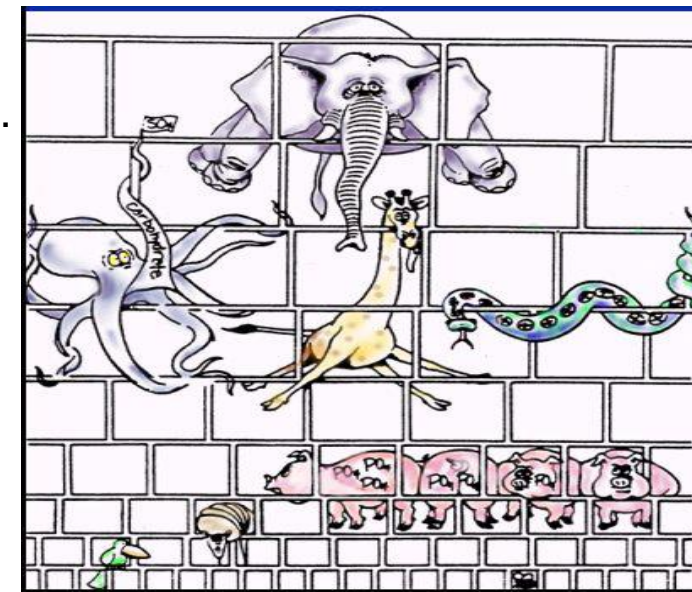
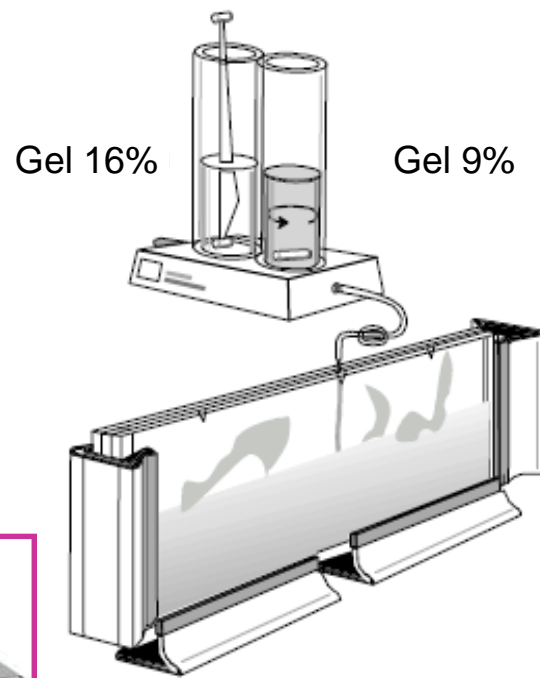
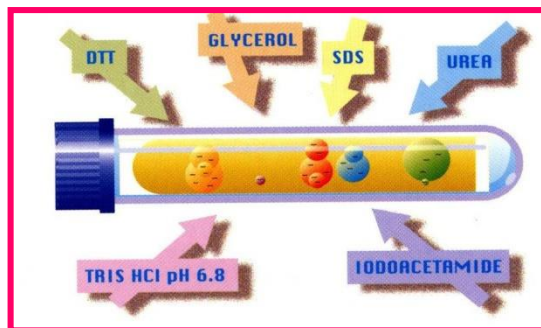
<sup>2</sup> This voltage may not be reached within the suggested step duration.

Si tratta di un gel di poliacrilammide nel quale la percentuale di acrilammide varia in modo uniforme dal 5% (alla sommità del gel) al 25% (sul fondo dello stesso).

Il gradiente è formato con un formatore di gradiente.

I gel in gradiente sono generalmente preparati con SDS e con uno *stacking* gel.

## Equilibratura dell'IPG strip



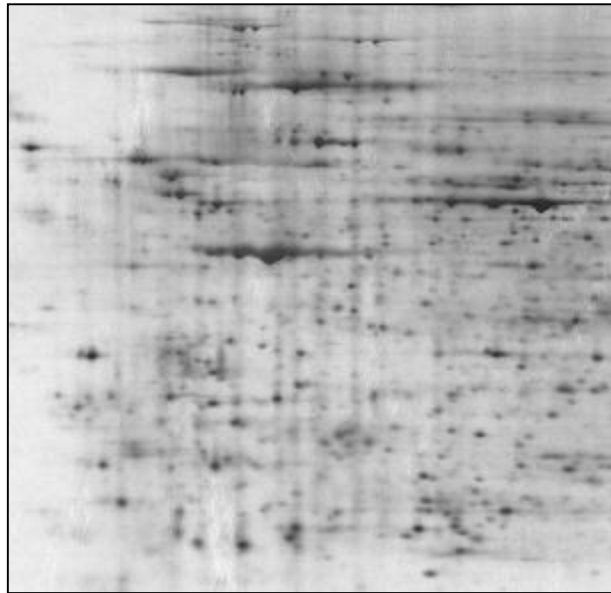


# Tecniche di rivelazione degli spot proteici

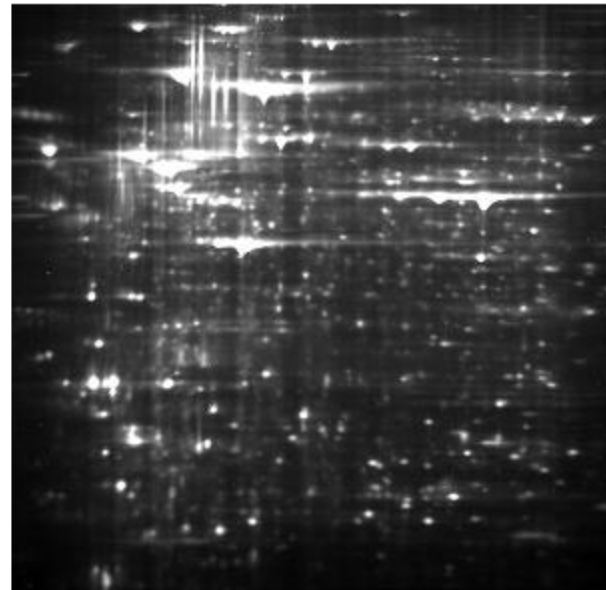
**COLORIMETRICHE**

COLORANTI ORGANICI

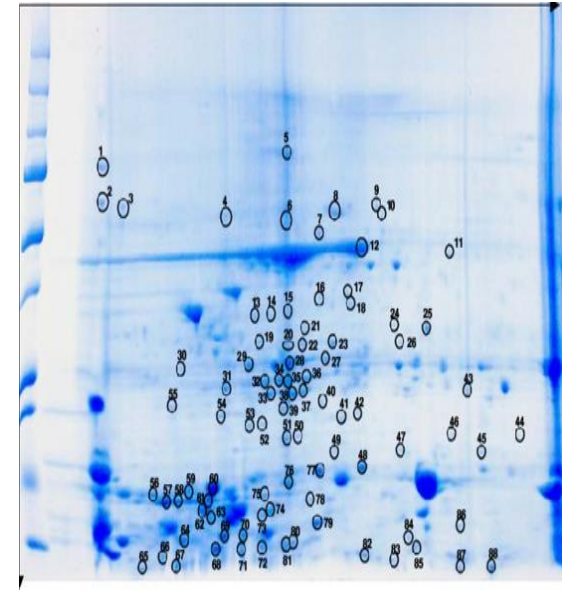
COLORAZIONE CON ARGENTO



Colorazione con silver stain



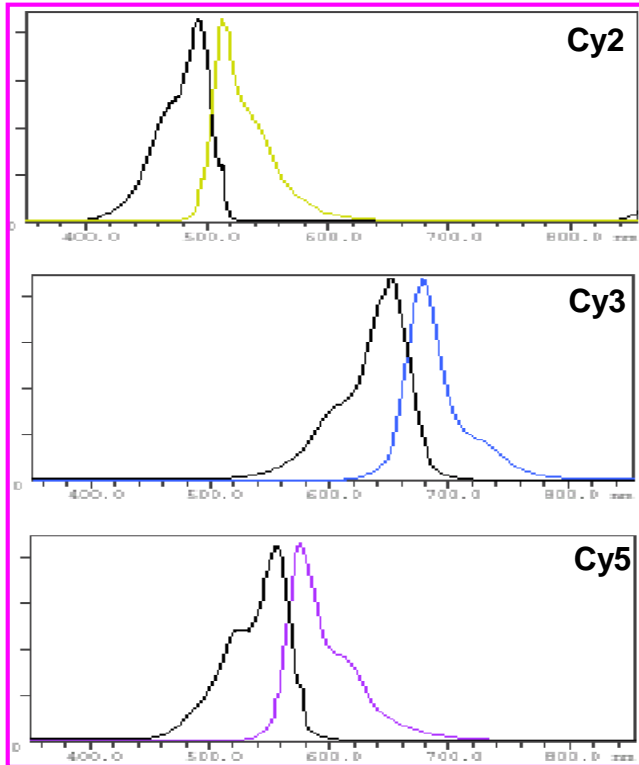
Colorazione con sypro ruby



Colorazione con blue Coomassie

# Tecniche di rivelazione degli spot proteici

Cy2, Cy3, Cy5 ⇒ esteri della cianina



**Fluorocromi**

Cy3: ecc: 532 - emiss: 580

Cy5: ecc: 633 - emiss: 670

Cy2: ecc: 488 - emiss: 520

## DIGE CyDye Fluors

### Minimal labeling dyes

Size and charge matched

3 colors: Cy2, Cy3, Cy5

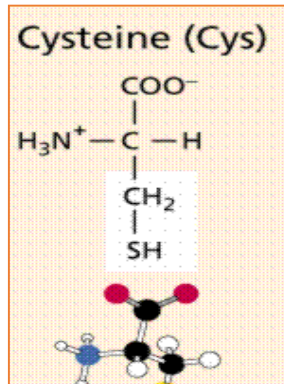
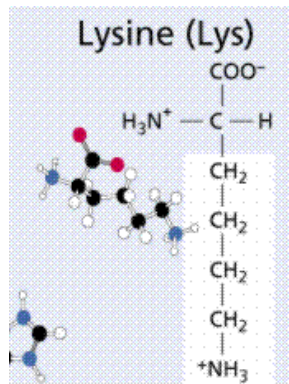
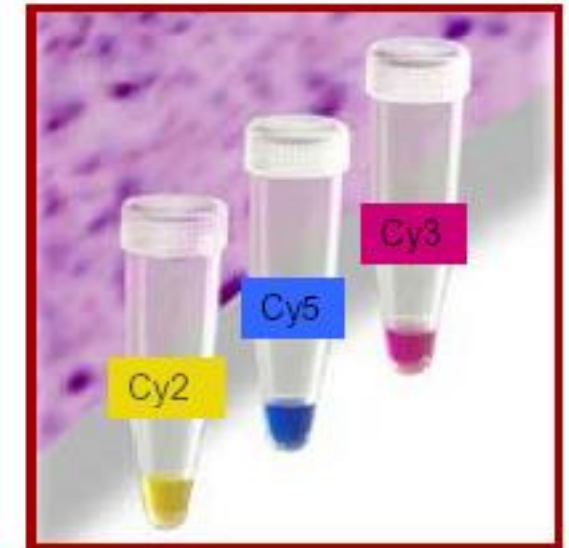
label ε-amino group of lysine

### Saturation labeling dyes

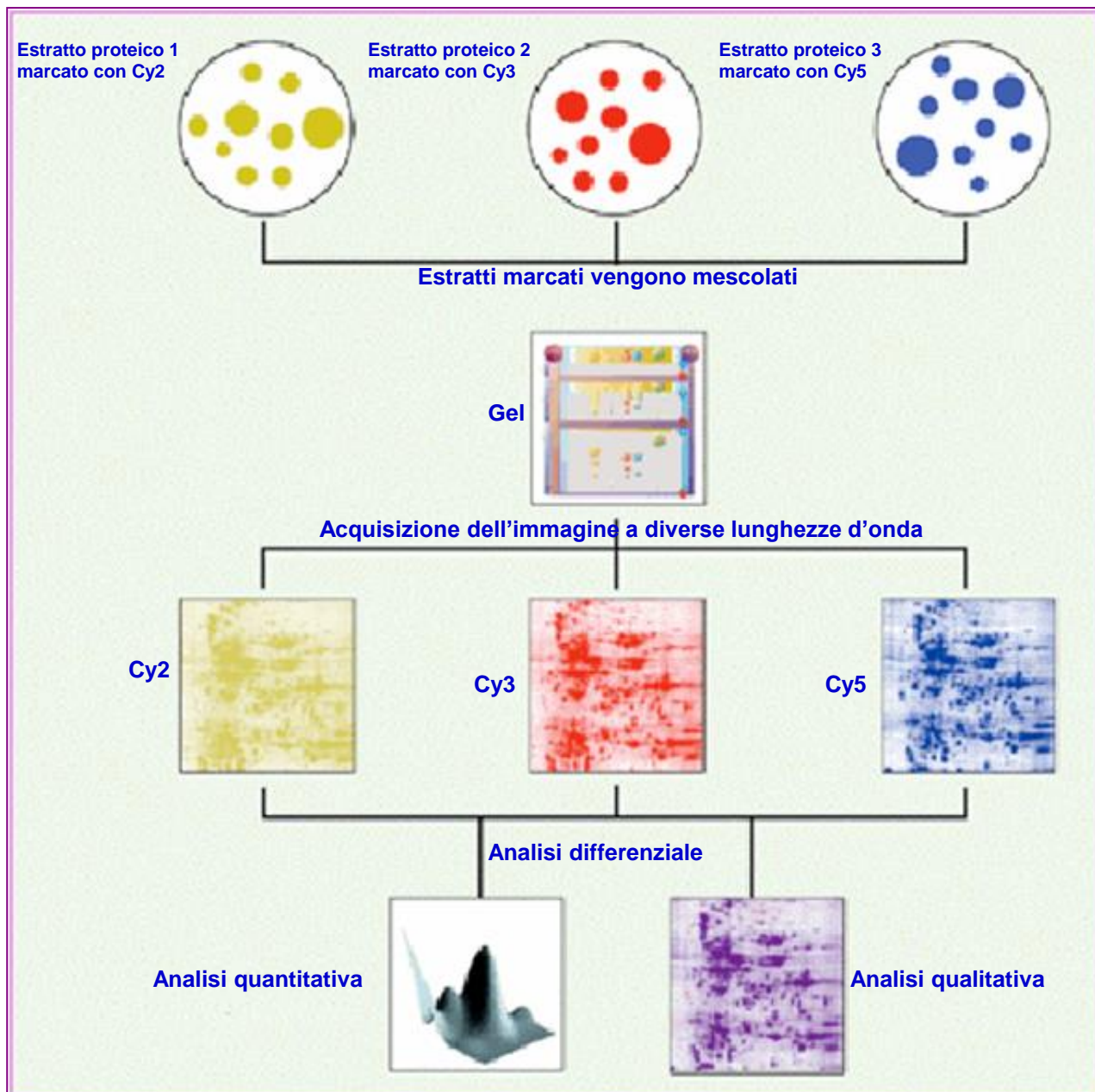
Size matched, neutral charge

2 colors: Cy3, Cy5

label thiol group of cysteine

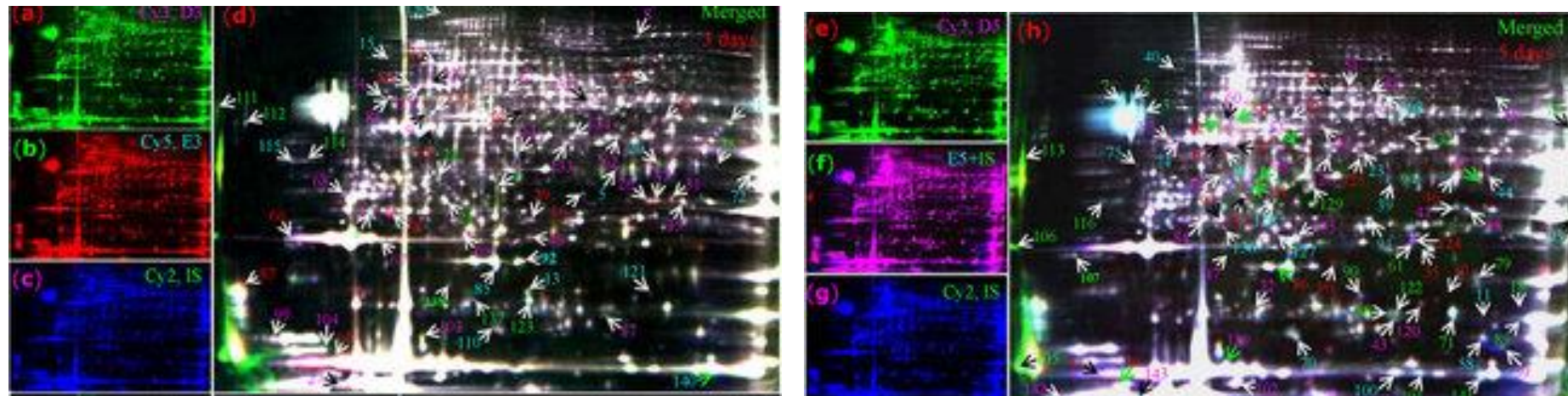


# Fluorescence 2-D Difference Gel Electrophoresis (DIGE)



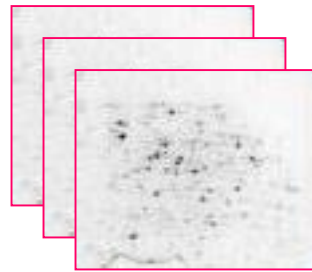
# Ettan DIGE system

1. Perché si sono sviluppate le scienze omiche



# Esempio di una tipica procedura di analisi proteomica

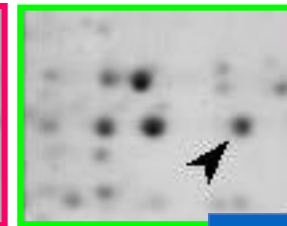
## 1 - Controllo



Controllo

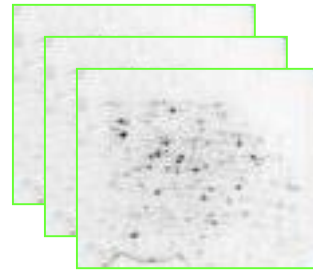
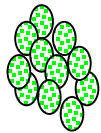


Trattato



Individuazione di una significativa differenza d'espressione

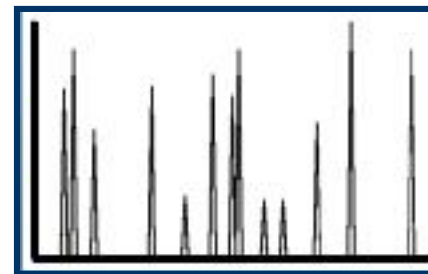
## 2 - Trattato



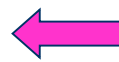
**Identificazione della proteina**



Taglio dello spot



Ricerca omologie in Banche dati



MALDI-TOF

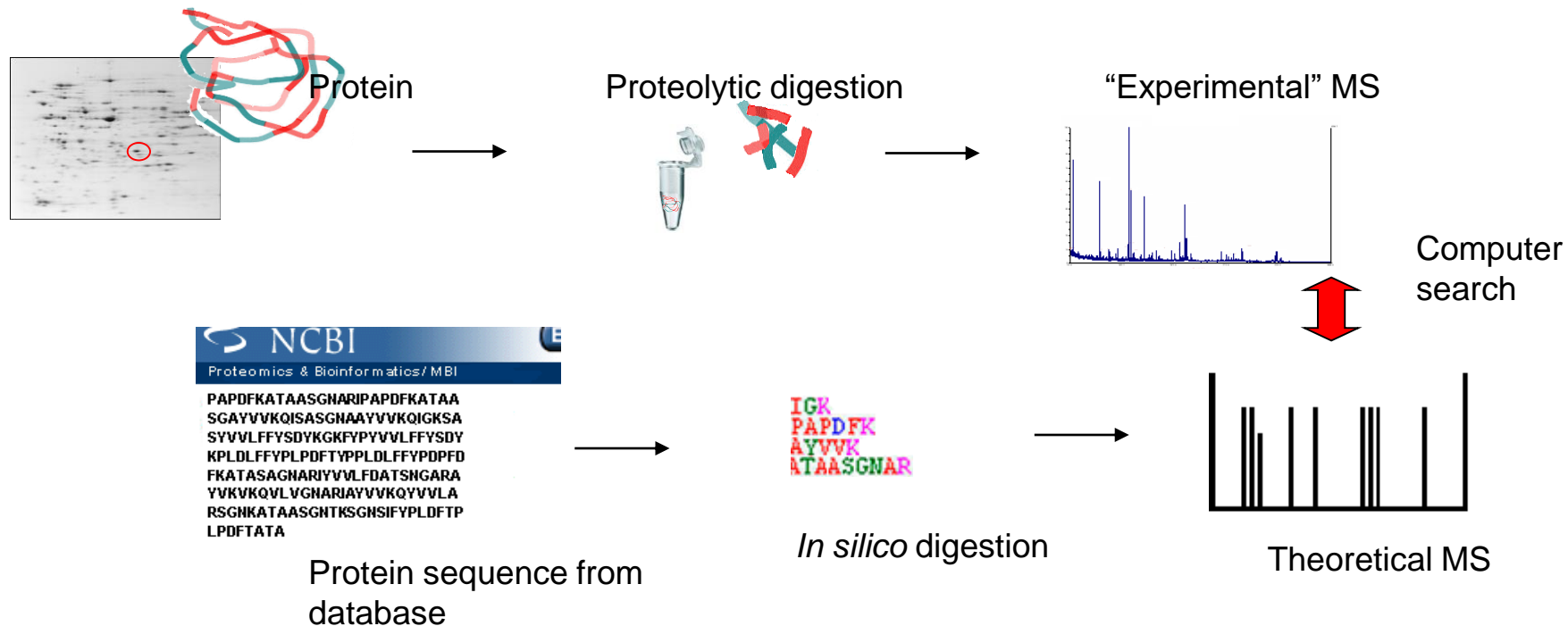
# Sequenziamento delle proteine mediante spettrometria di massa MALDI-TOF

A protein identification technique, that correlates experimental data with theoretical data.

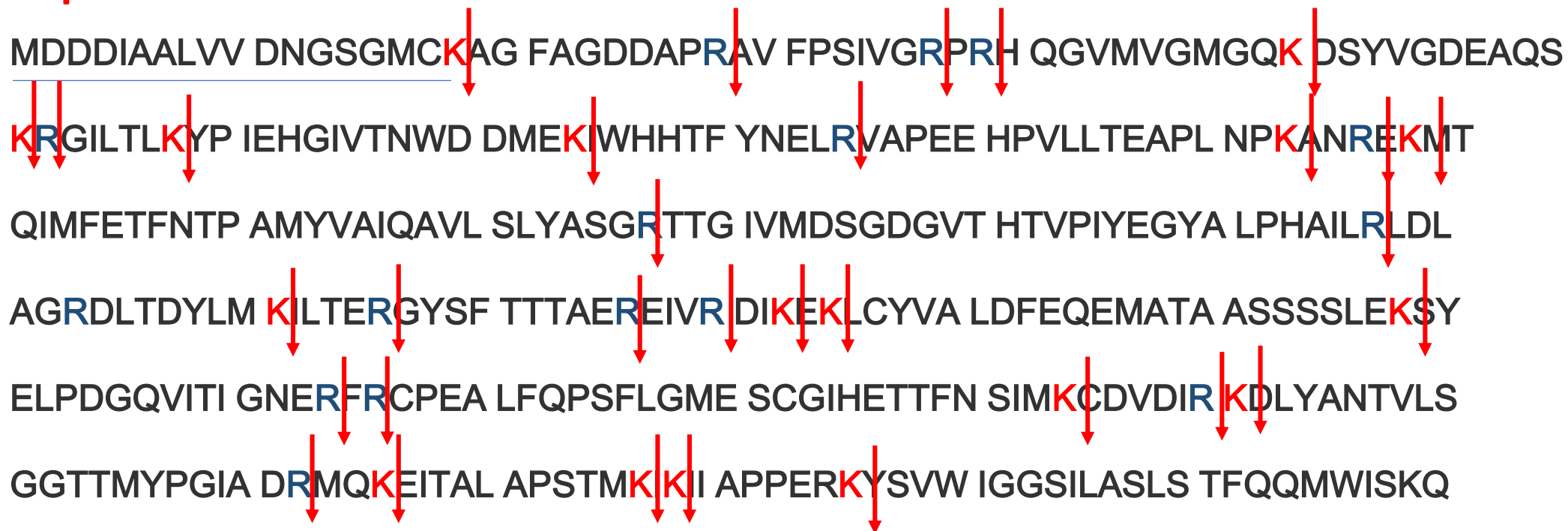
1. SPOT PICKING
2. DECOLORAZIONE
3. DIGESTIONE ENZIMATICA
4. ESTRAZIONE DEI PEPTIDI
5. CONCENTRAZIONE/PURIFICAZIONE
6. ANALISI MALDI-TOF

*Success depends upon:*

- *Avoiding contamination of samples*
- *Digesting the protein efficiently*
- *Maximizing recovery of peptides*
- *Minimizing losses from handling*



# La tripsina bovina purificata viene usata per digerire le proteine in provetta ed ottenere dei peptidi, la cui massa molecolare viene accuratamente misurata dallo spettrometro di massa



## Atomic composition:

Carbon	C	1850
Hydrogen	H	2903
Nitrogen	N	491
Oxygen	O	561
Sulfur	S	23

Formula:  $C_{1850}H_{2903}N_{491}O_{561}S_{23}$

Total number of atoms: 5828

## Massa monoisotopica

MDDIAALVVDNGSGMCK 1853.8078

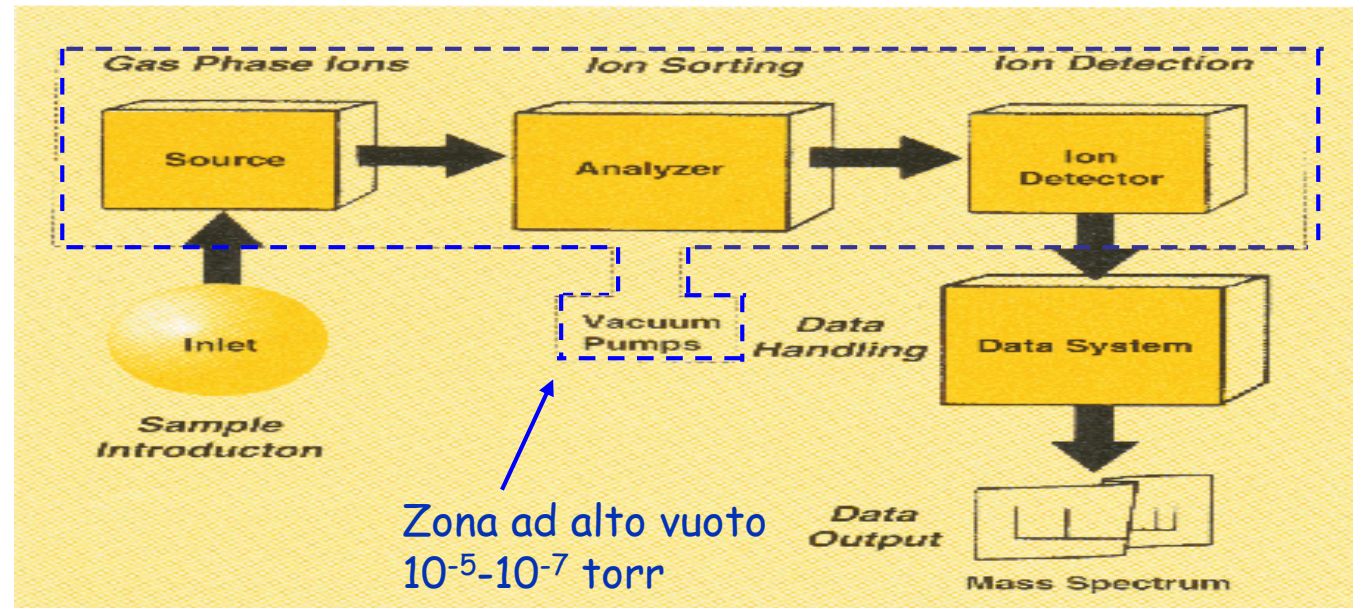
Proteolytic digestion



# Spettrometria di massa

Lo spettrometro di massa è uno strumento che produce ioni e li separa in fase gassosa in base al loro rapporto massa/carica ( $m/z$ ). Quando una molecola, in fase vapore, viene investita da un fascio di elettroni dotati di notevole energia cinetica ionizza a ione monopositivo. Lo ione, o l'insieme di ioni, vengono separati sfruttando un campo elettrico e/o magnetico.

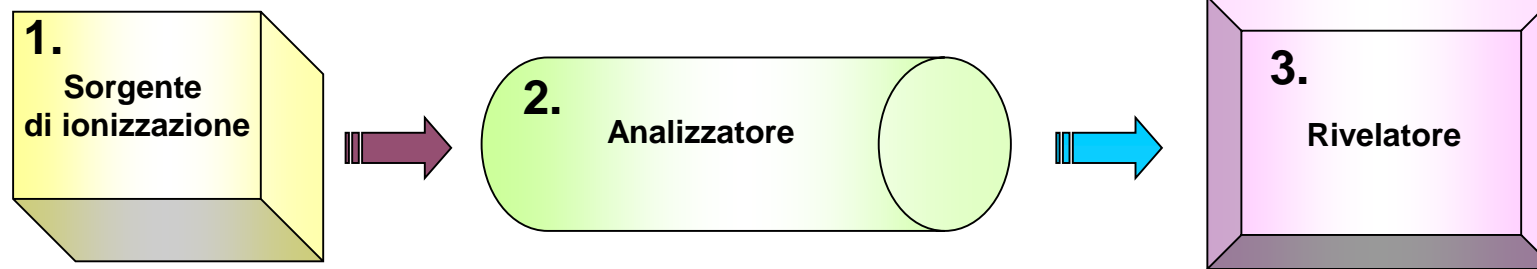
**Ionizzazione:** processo per cui una molecola elettricamente neutra, sottoposta ad un flusso energetico, perde un elettrone e si carica positivamente.



- camera di introduzione del campione.
- sorgente di ioni: un flusso di elettroni viene generato da un filamento incandescente ed accelerato verso la camera di ionizzazione
- analizzatore delle masse ioniche: il raggio di ioni creato nella camera di ionizzazione viene separato per mezzo del passaggio attraverso un campo elettrico e/o magnetico.
- Collettore e rivelatore ionico: i raggi ionici separati vengono raccolti
- Analisi dello spettro di massa: Diagramma che riporta l'abbondanza relativa degli ioni in funzione del rapporto  $m/z$ .



# Differenti tipi di spettrometria di massa



## 1. Tipi di ionizzazione

### Hard Ionization:

1. Ionizzazione elettronica
  - Electron Ionization (EI)

### Soft Ionization:

- Ionizzazione chimica
  - Chemical Ionization (CI)
- Bombardamento con atomi veloci
  - Fast Atom Bombardment (FAB)
- Ionizzazione termospray
  - Thermospray Ionization (TSP)
- Ionizzazione elettrospray
  - Electrospray Ionization (ESI)
- Ionizzazione chimica a pressione atmosferica
  - Atmospheric Pressure Chemical Ionization (APCI)
- **Desorbimento laser assistito da matrice**
  - **Matrix Assisted Laser Desorption (MALDI)**

## 2. Tipi di analizzatori

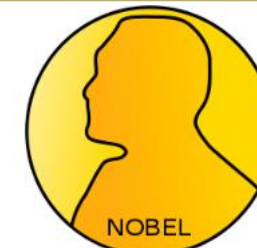
- Analizzatori magnetici
- Analizzatori a quadrupolo
- Analizzatori a trappola ionica
- **Analizzatori TOF (time of flight)**

## 3. Tipi di Rivelatori

- Faraday cup
- Moltiplicatore elettronico

2002 Chimica

Premio Nobel



Alfred Nobel

	John B. Fenn
	Koichi Tanaka
	Kurt Wüthrich

## Matrix Assisted Laser Desorption (MALDI):

Permette di ottenere ioni in fase gassosa tramite pulsazioni provenienti da un laser (generalmente ad azoto,  $\lambda = 337 \text{ nm}$ ).

Il composto da analizzare è mescolato con una matrice (che ha massimo di assorbimento attorno a 337 nm) che cristallizzano su un supporto inerte).

Le pulsazioni laser causano vaporizzazione e ionizzazione per trasferimento di un protone dalla matrice al campione

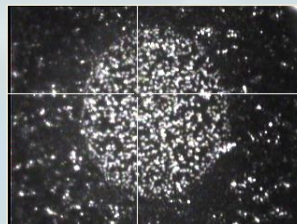
### Analizzatore TOF

Separa gli ioni aventi differente rapporto carica/massa ( $m/z$ ). Il tempo che lo ione impiega ad attraversare il tubo di volo è maggiore per gli ioni più pesanti e minore per quelli più leggeri.

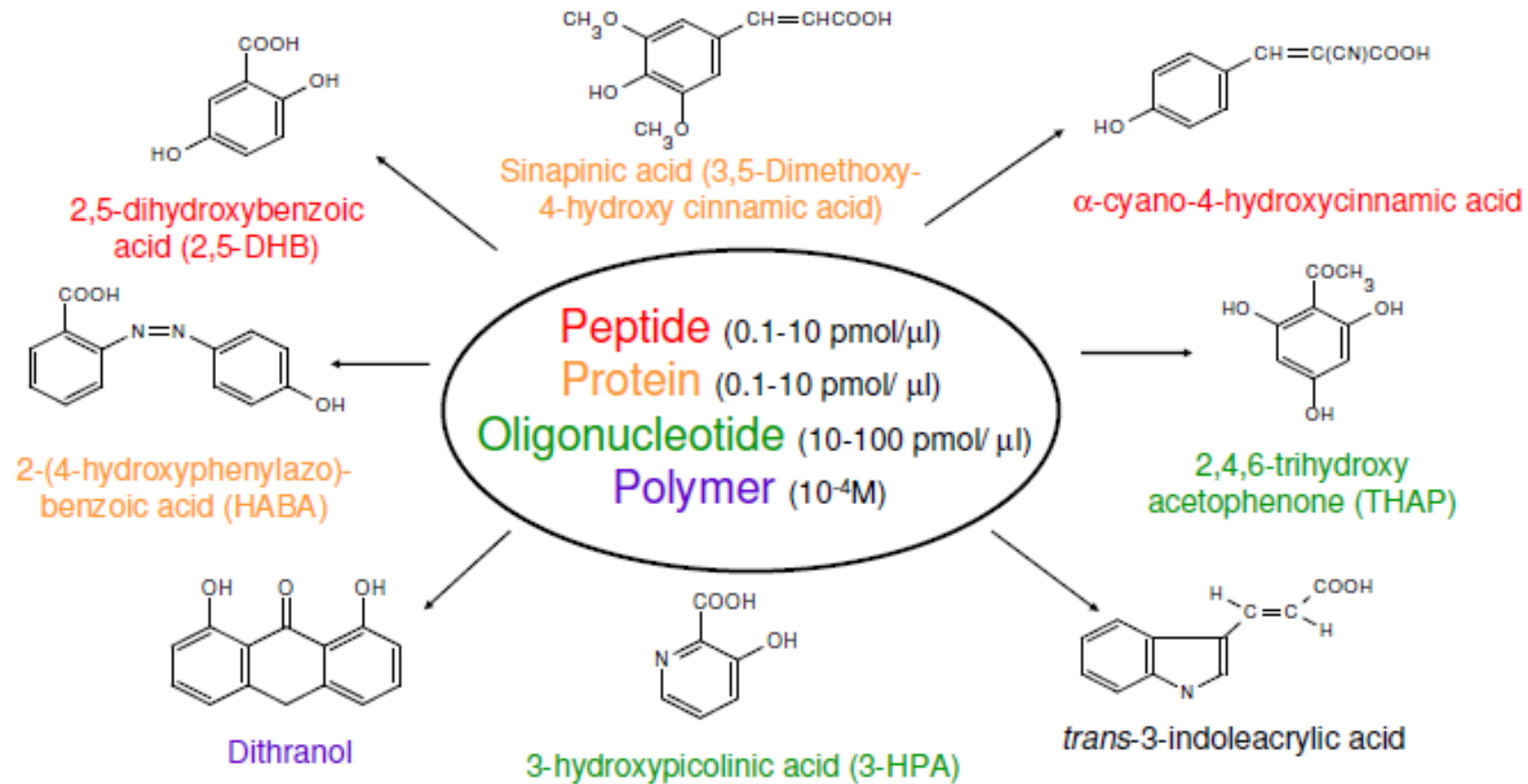
Nel sistema MALDI il campione da analizzare è disciolto in un solvente volatile contenente piccole molecole organiche (acidi aromatici che rappresentano la matrice, il cui gruppo aromatico per l'appunto ha un massimo di assorbimento alla  $\lambda$  del raggio laser, mentre la parte acida ionizza il campione). La miscela analita-matrice viene fatta asciugare su un supporto di acciaio per favorire la cristallizzazione dei peptidi con la matrice

#### MALDI preparation / MS analysis:

- 1.) **Mix sample and matrix** solutions in suitable ratio.
- 2.) Put a sub  $\mu\text{l}$  aliquot of this mixture on the **target plate**.
- 3.) Let the mixture **co-crystallize**.
- 4.) Insert the target plate in the MALDI mass spec (running under **high vacuum**).
- 5.) **Shoot with the laser** on the MALDI preparation to **generate ions** (**singly charged** predominantly).
- 6.) **Collect, analyze** and **detect** the resulting **ions**.



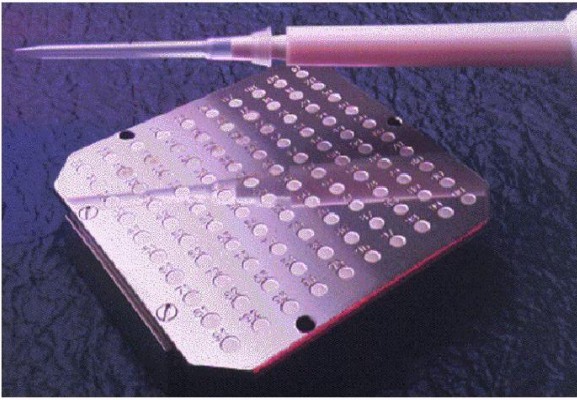
# Matrici



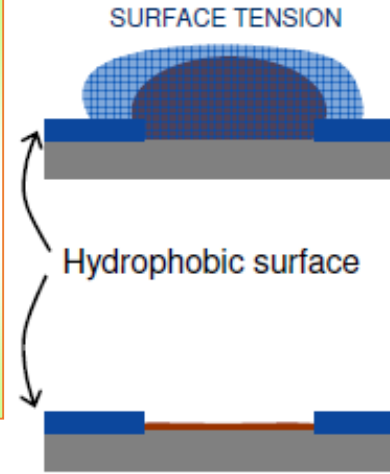
## Caratteristiche

- forte assorbimento alla lunghezza d'onda del laser
- Stabilità a basse pressioni (sottovuoto)
- Capacità di promuovere la ionizzazione dell'analita
- Solubilità nei solventi compatibili con l'analita
- Assenza di reattività chimica nei confronti del campione

# Preparazione del campione

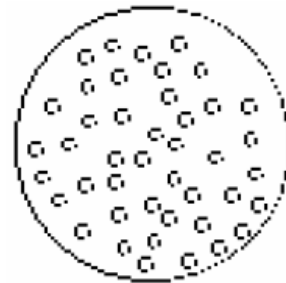


- Procedure for Plate Washing
- Clean the plate with non-ionic detergent
  - Rinse with abundant tap water
  - Place pure ACN over the surface and let it stay for a few seconds (**for stain steel plates only**). Repeat 2-3 times
  - Pour 50% ACN on the plate and let it stay for a few seconds. Repeat 2-3 times
  - Rinse with distilled water
  - Air dry



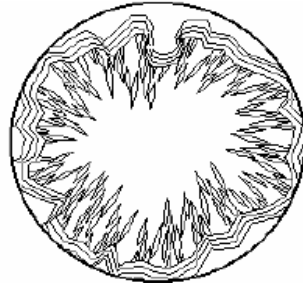
# Preparazione della matrice e cristallizzazione

© 2005 Applied Biosystems



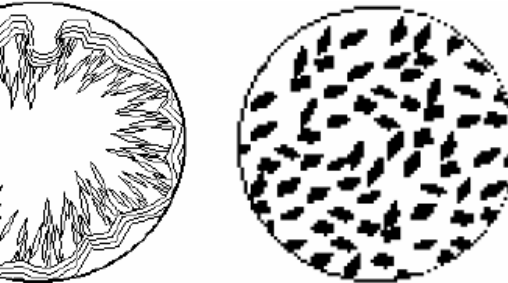
$\alpha$ -cyano

5 mg/ml in  
50% ACN  
0,1% TFA



DHB

10 mg/ml in  
water or 50%  
ACN



Super-DHB

A = 10 mg/ml DHB  
in 20% ACN  
B = 10 mg/ml 5-  
methoxysalicylic acid  
in 50% ACN  
proteins, large peptides  
Combine A:B (9:1)

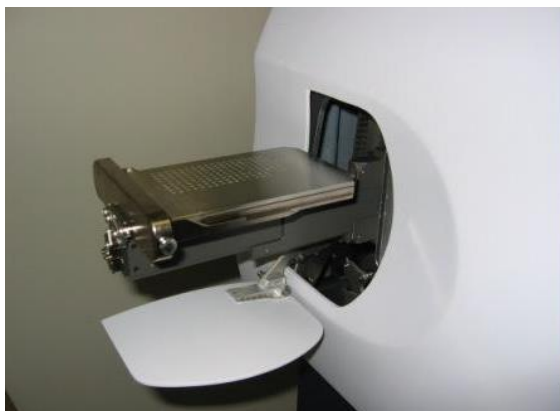
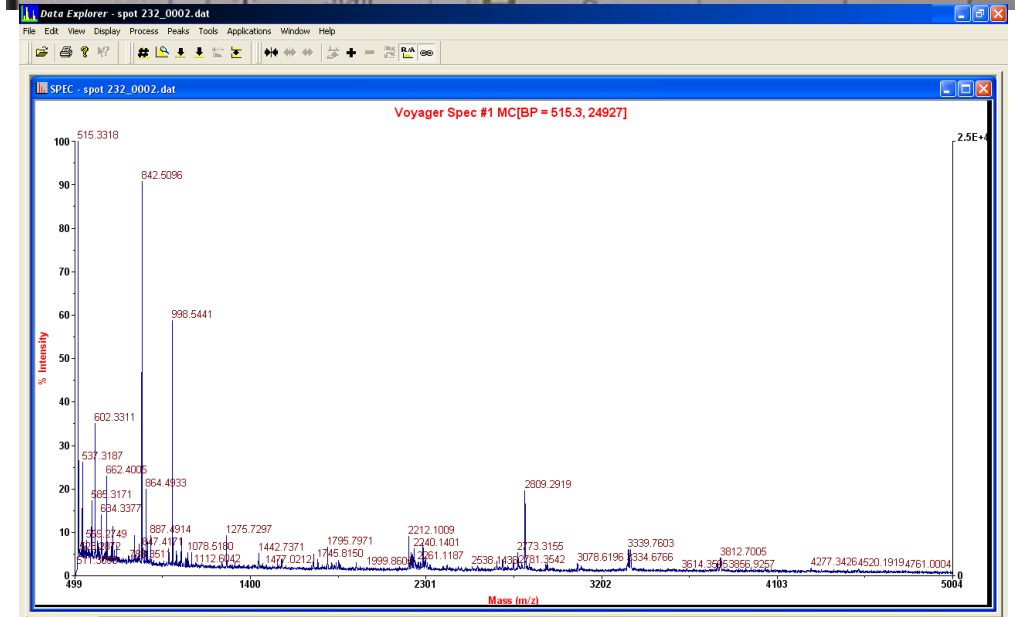
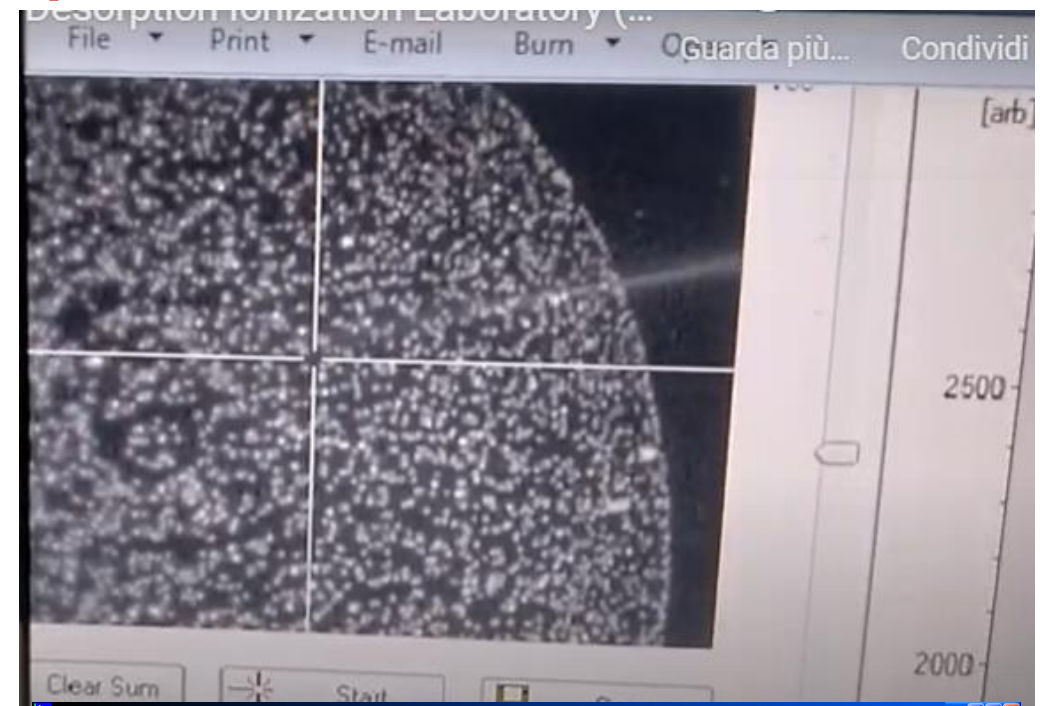
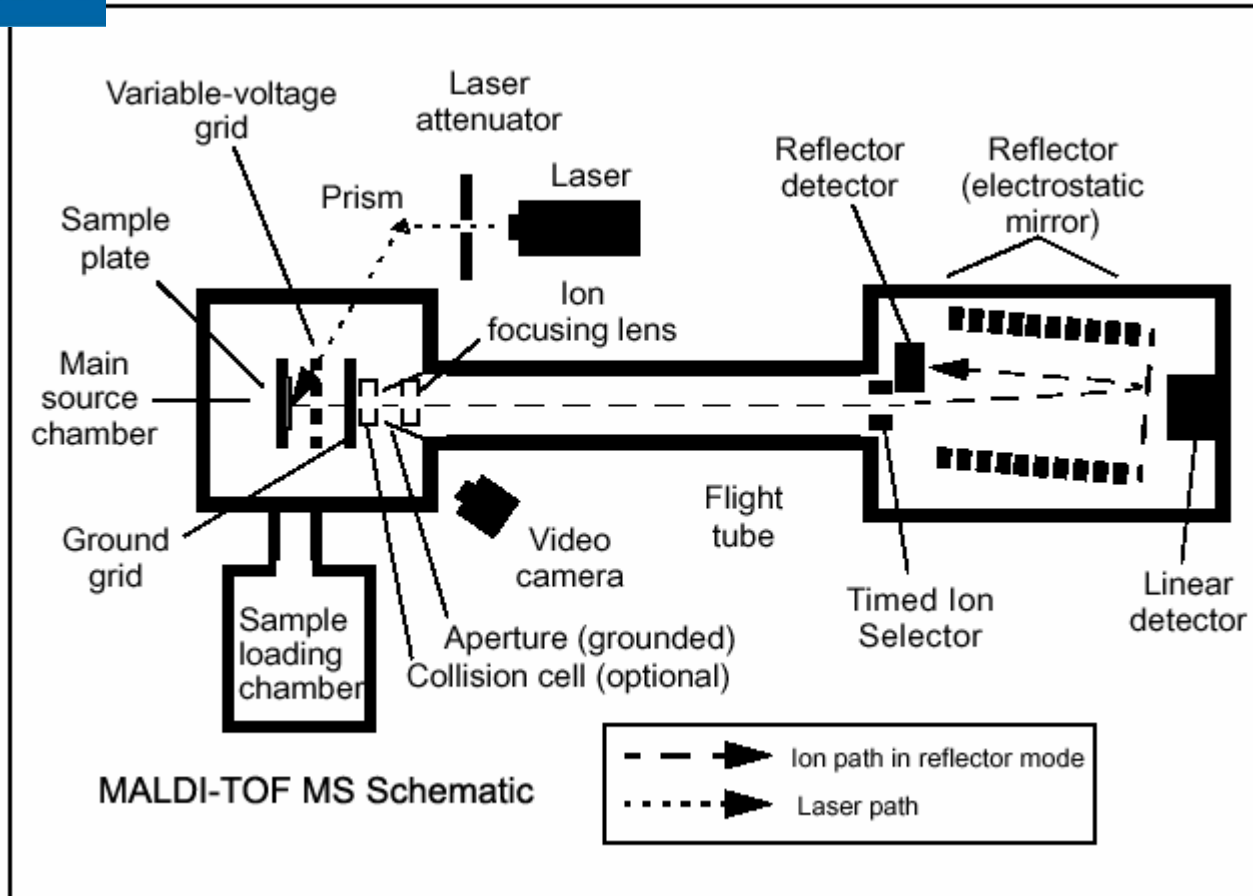
Sinapinic acid

10 g/L in 30-50%  
ACN with 0.1%  
TFA

peptides, protein  
digests, small proteins  
and PNAs below 10,000  
Da

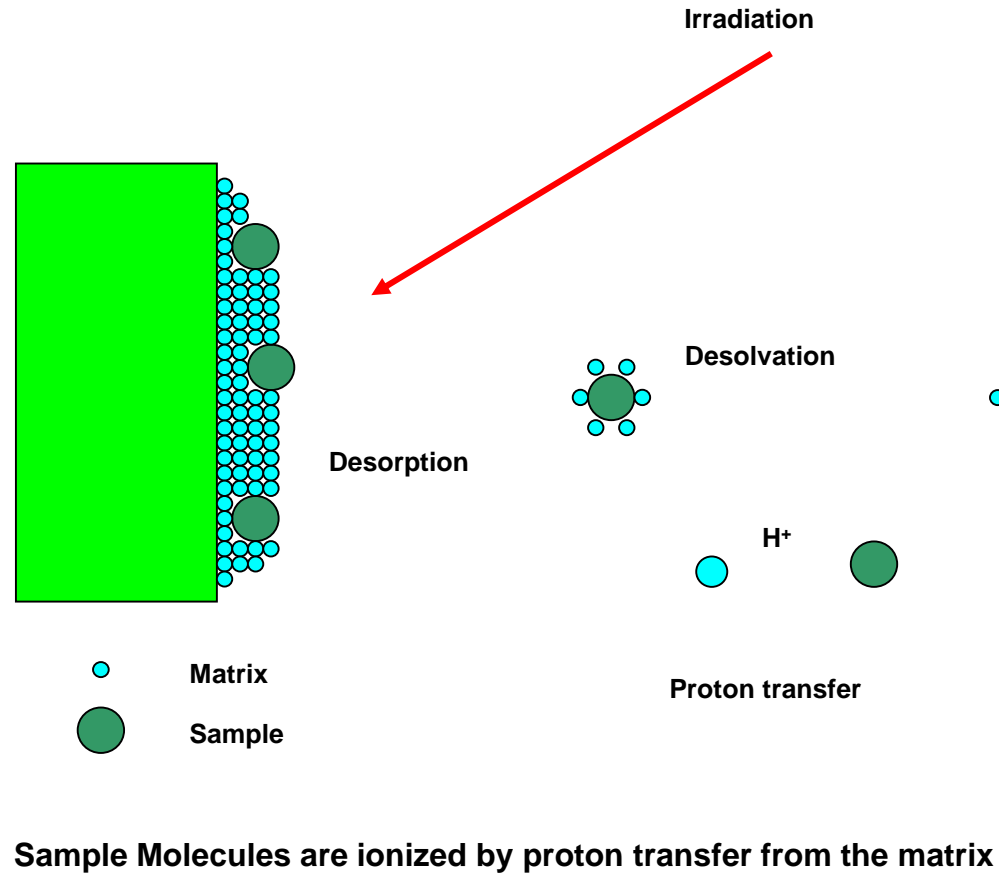
small molecules, phosphopeptides, peptides, neutral carbohydrates,  
some synthetic polymers

# Introduzione del campione



Vi sono tre differenti camere che si aprono sequenzialmente e sono collegate ad un sistema di pompaggio responsabile della creazione del vuoto

# Sorgente ionica Matrix Assisted Laser Desorption (MALDI):



Si basa sul processo di desorbimento/ionizzazione effettuato mediante l'interazione tra il raggio laser pulsante (azoto) e l'analita solido (peptidomatrice)

Le intense e brevi pulsazioni del laser causano il riscaldamento e la sublimazione dei cristalli di matrice, con conseguente espansione in fase gassosa delle molecole di matrice-analita. Il trasferimento protonico dalla matrice all'analita determina la ionizzazione e la presenza di un campo elettrostatico induce l'accelerazione degli ioni prodotti verso l'analizzatore

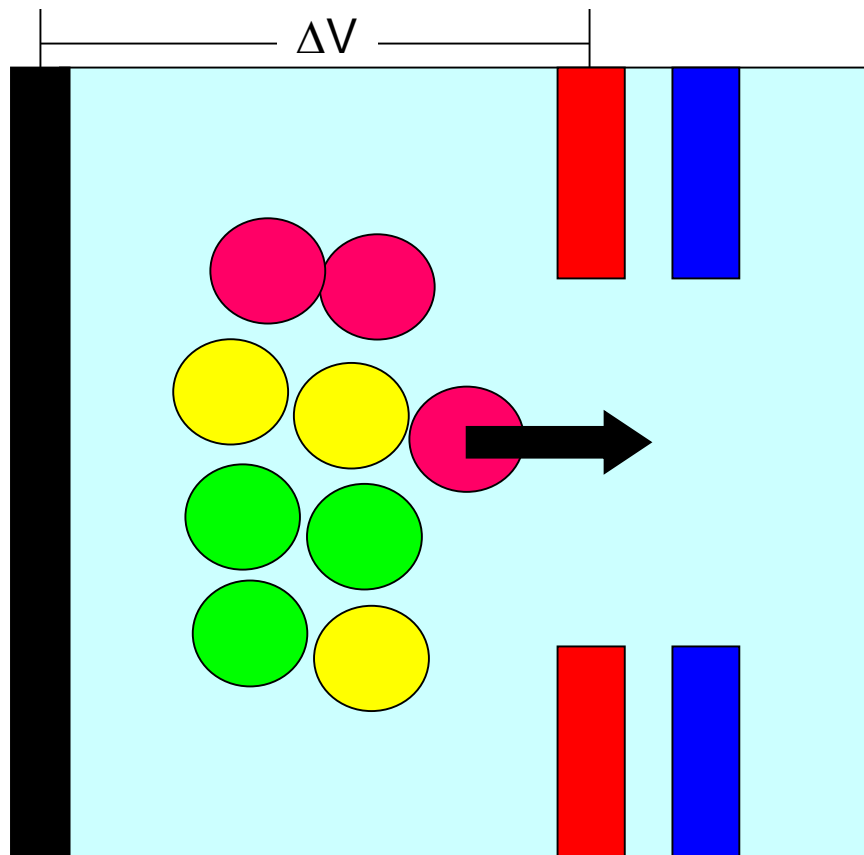
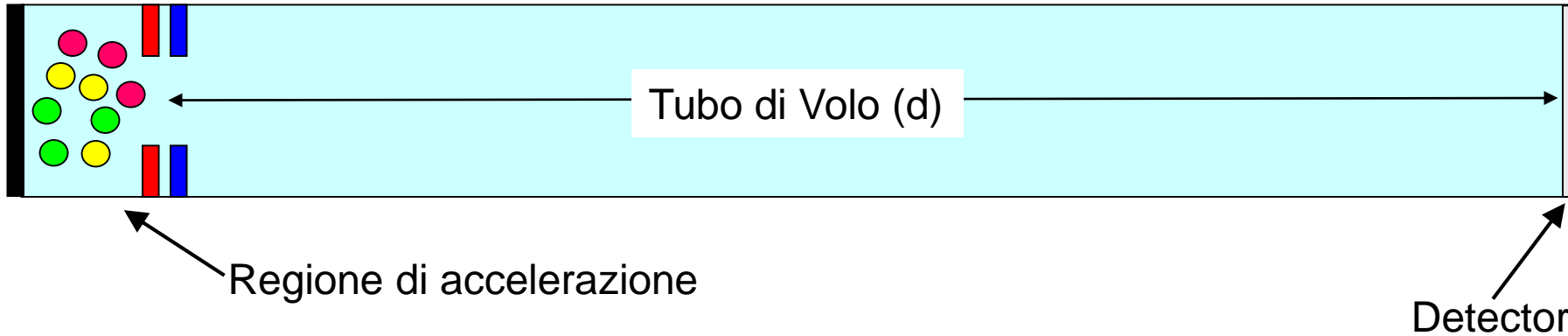
## Caratteristiche del laser:

**Fluenza:** energia fornita per unità di area (10-100 mJ/cm<sup>2</sup>)

**Irradianza:** Fluenza diviso la durata dell'impulso;

**Durata dell'impulso:** I laser UV emettono impulsi che hanno una durata di 3-10 ns, cioè abbastanza breve per garantire che si abbia degradazione termica del campione e per aumentare la risoluzione

# Analizzatore TOF: (Time of Flight)



Nella regione di accelerazione, gli ioni vengono sottoposti ad una forte differenza di potenziale che impartisce ad ognuno degli ioni una propria velocità

$$mv^2/2 = zeE$$

Dal momento che  $v=s/t$  e  $s=d$

Si ha che  $m(d/t)^2 / 2 = zeE$

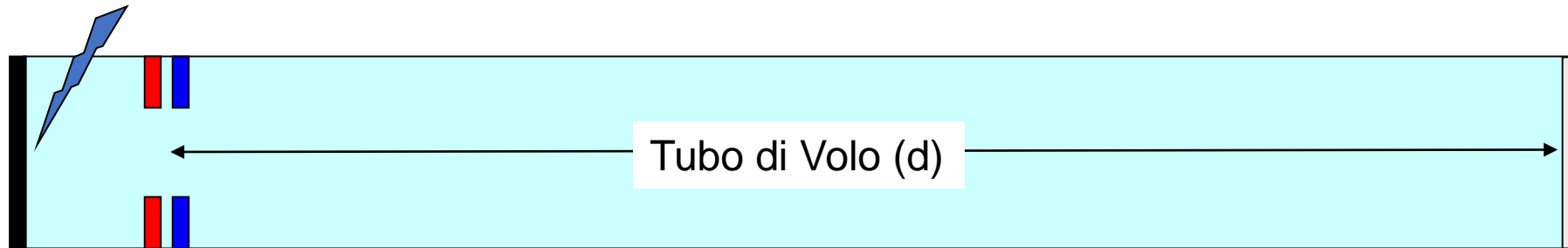
$$md^2/t^2 = 2zeE$$

=>

$$t^2 = \frac{m}{z} (d^2/2eE)$$

**Tempo** che lo ione impiega per percorrere il tubo di volo e **dipendente** dal suo rapporto  **$m/z$**

Generazione degli ioni – Ad esempio tramite un impulso laser (MALDI)

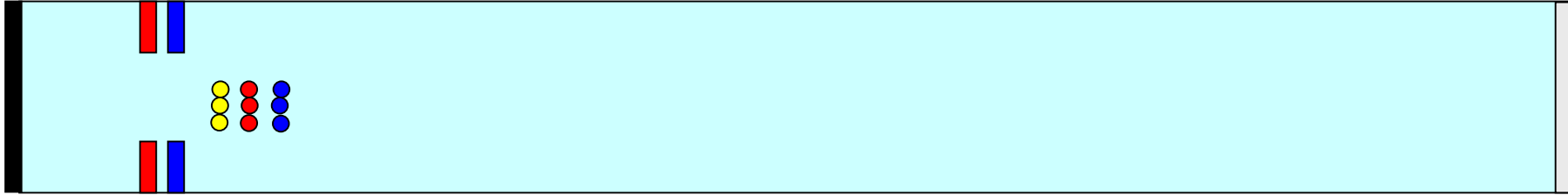




Accelerazione degli ioni mediante  $\Delta V$



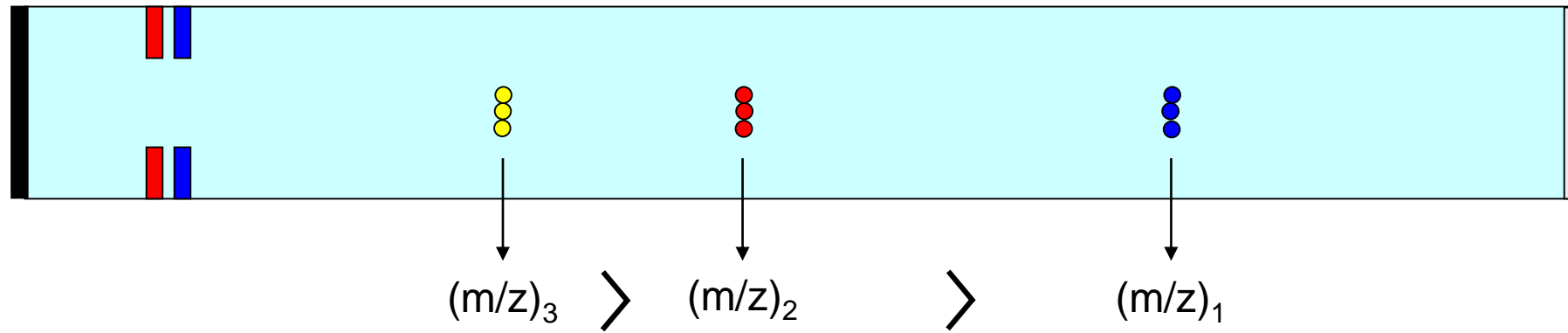
Ioni hanno acquisito una propria velocità  
in dipendenza del loro  $m/z$  e si muovono  
verso il detector



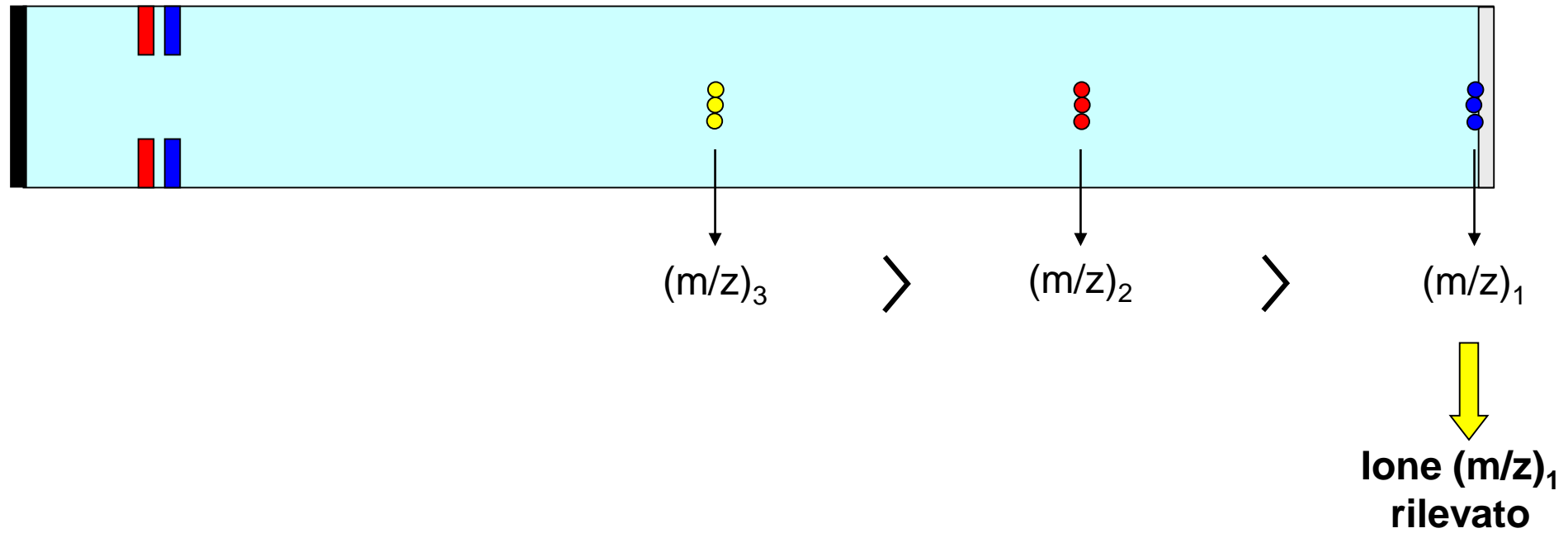
Le loro diverse velocità separano progressivamente gli ioni con diverso  $m/z$  durante il loro cammino verso il detector



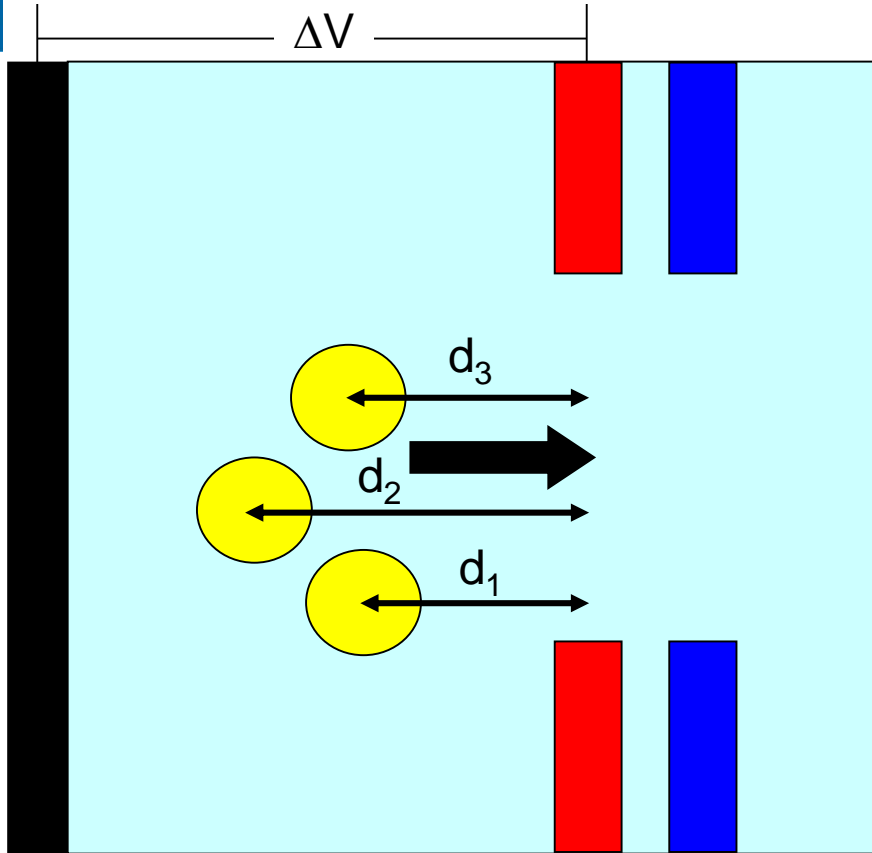
Le loro diverse velocità separano progressivamente gli ioni con diverso  $m/z$  durante il loro cammino verso il detector



Gli ioni con  $m/z$  diversi raggiungono il detector in tempi diversi



Questo in teoria...ma in pratica?



Gli ioni prodotti in sorgente non partono tutti dalla medesima posizione

=>

Sono soggetti ad accelerazione differenti

=>

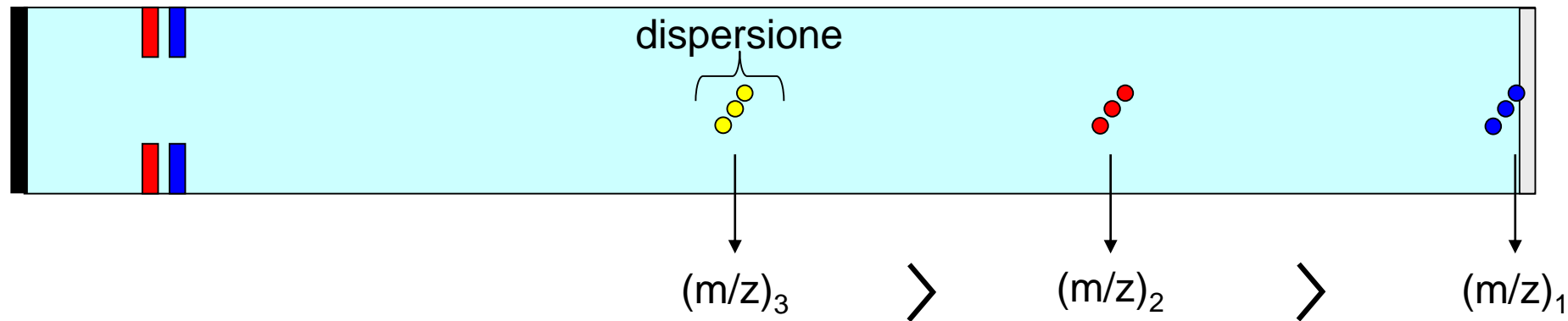
Si muovono con velocità differenti

=>

Pur avendo  $m/z$  uguale, raggiungono il detector in tempi diversi

=>

Risoluzione strumentale bassa

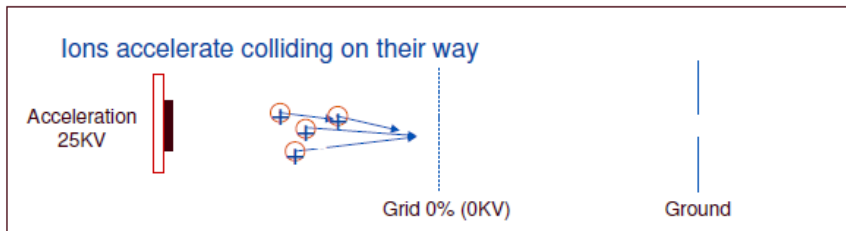
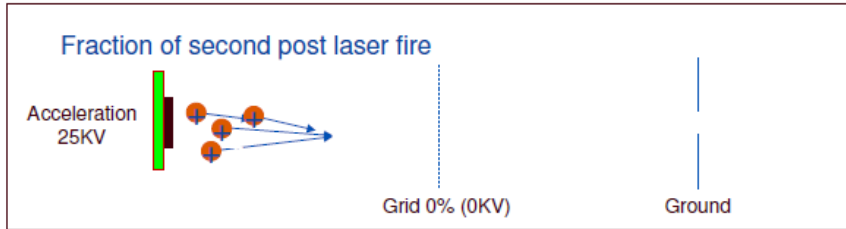


# Delayed extraction Technology

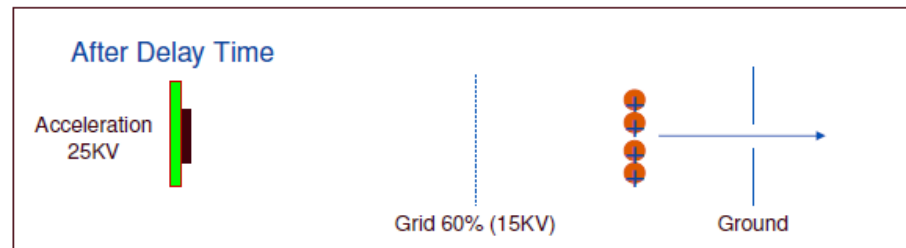
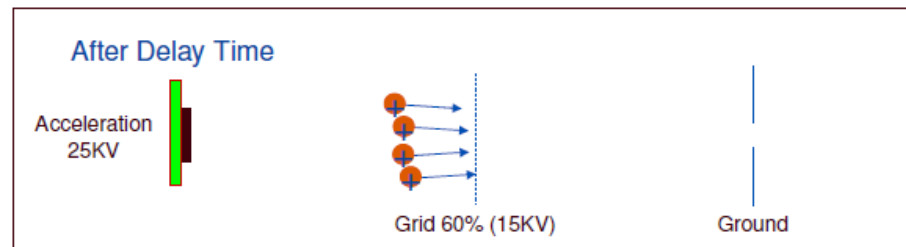
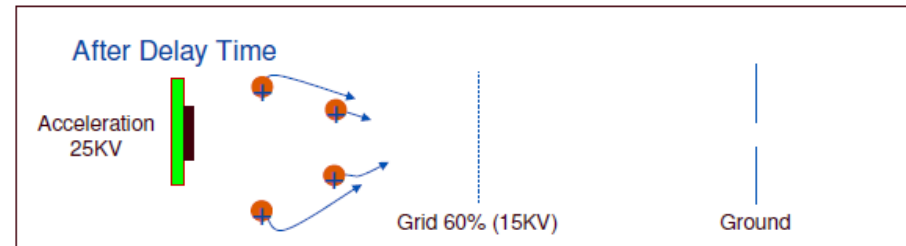
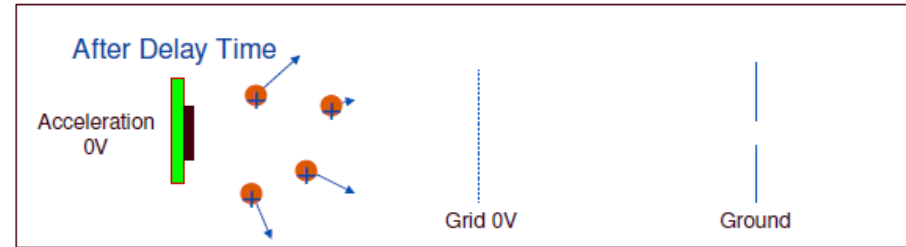
Applicazione di un tempo di ritardo tra la formazione degli ioni e la loro estrazione dalla sorgente.

Gli ioni prodotti in sorgente rimangono intrappolati tra la piastrina ed una griglia addizionale, le quali sono tenute ad una differenza di potenziale prossima a zero per un tempo dell'ordine dei nano-secondi. In tale periodo la collisione tra i vari ioni crea una distribuzione energetica iniziale più omogenea.

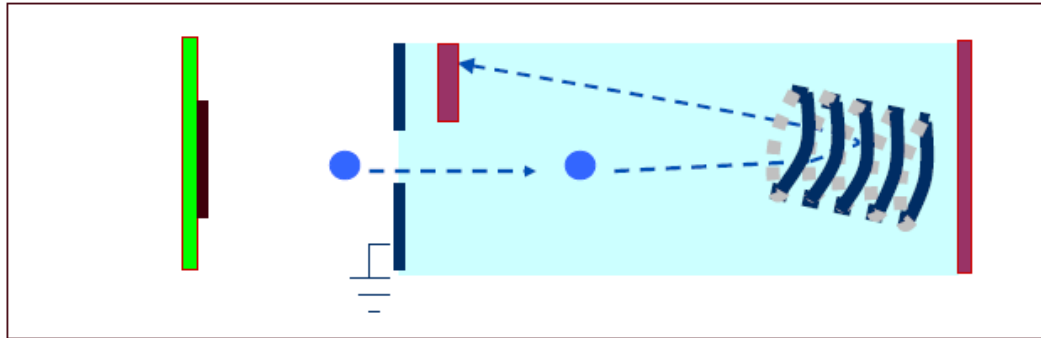
## No Delayed extraction



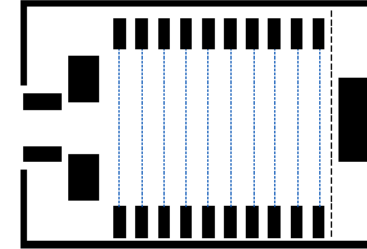
## Delayed extraction



# Reflector

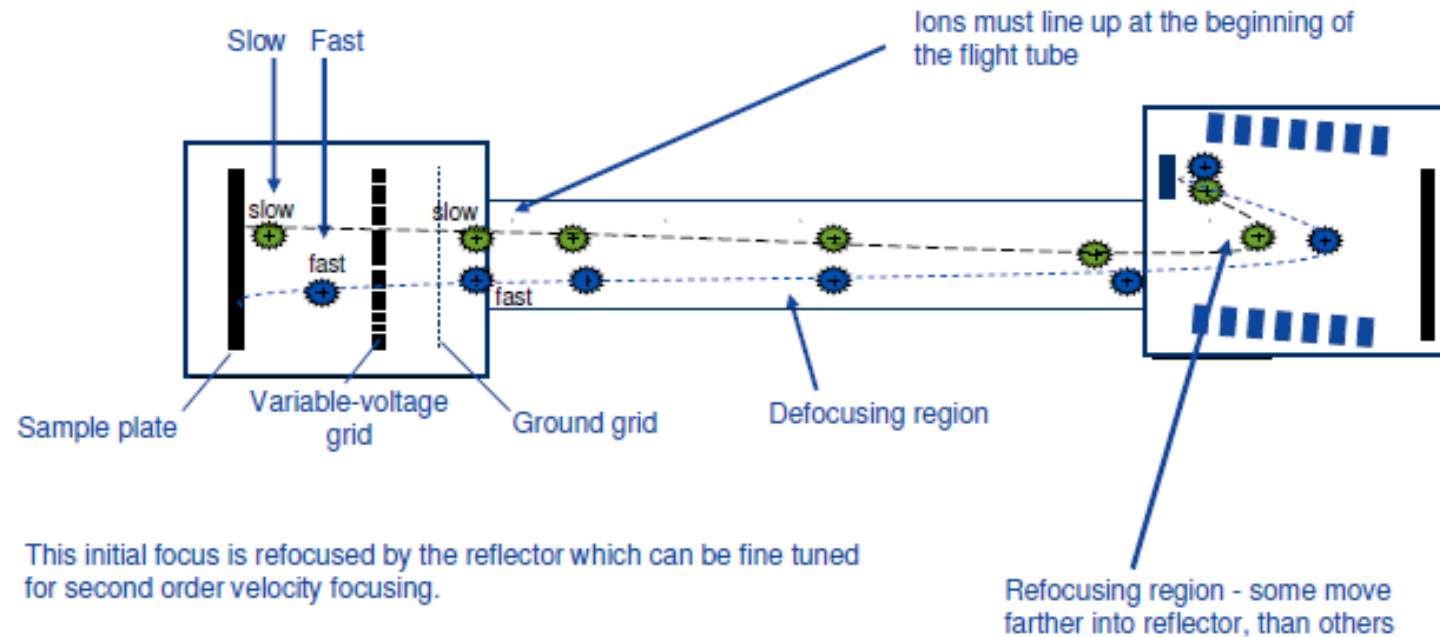


- The reflector is an electrical mirror with a voltage potential applied across the sides.
- The ions are sequentially slowed down through the reflector



Il reflector migliora la risoluzione in 2 modi:

- Incrementa l'effettiva lunghezza del tempo di volo
- Ri-focalizza ioni aventi stesso rapporto massa/carica inizialmente dispersi dai fenomeni correlati alla ionizzazione





# Parametri da regolare per l'acquisizione

**Range di massa:** 500-5000Da, dimensione attese per frammenti triptici

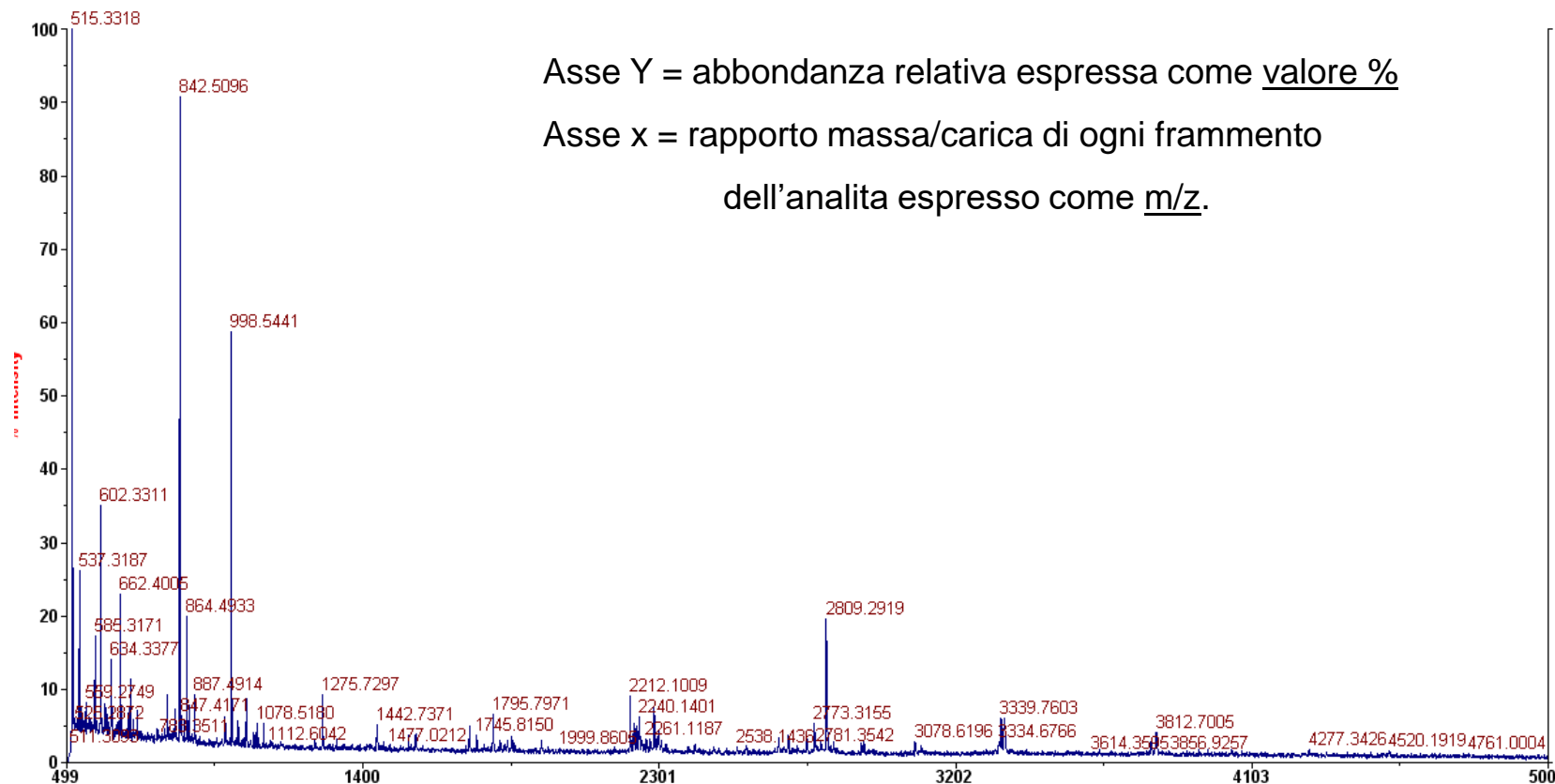
**Numero di pulsazioni laser incidenti sulla matrice:** 100/s

**Tipo di analizzatore:** lineare-reflecter

**Voltaggi** delle griglie responsabili dell'accelerazione edella  
sincronizzazione degli ioni desorbiti

## Spettro di massa

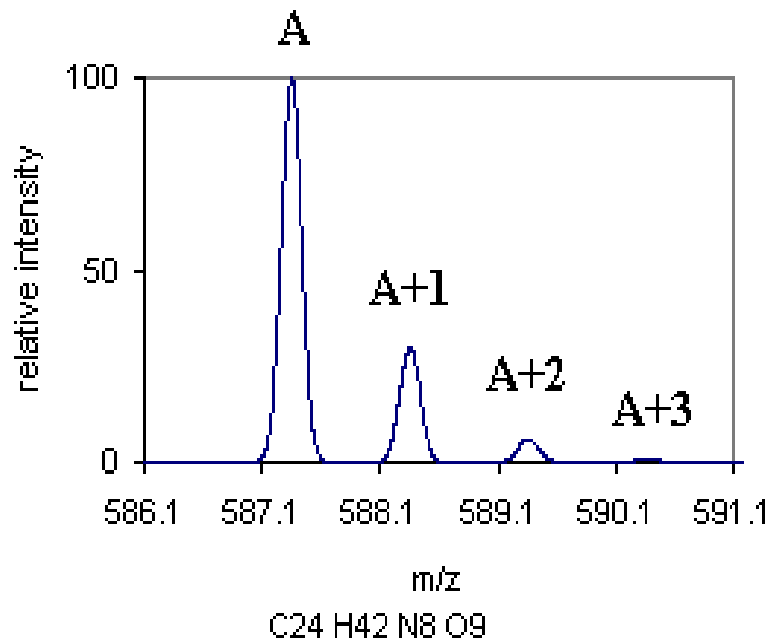
Voyager Spec #1 MC[BP = 515.3, 24927]



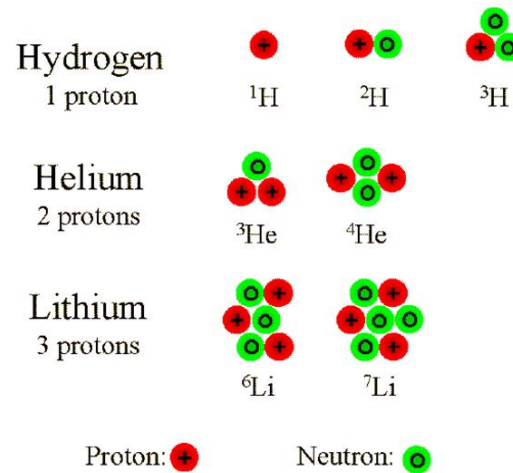
# Lo spettro di massa e gli isotopi

**Isotopi = differenti forme di un elemento, che hanno lo stesso numero di protoni (stesso numero atomico), ma differente numero di neutroni (massa atomica differente)**

Elements that are found in nature in form of only one single isotope, are called monoisotopic elements.



**A basse masse il picco monoisotopico è sempre più abbondante;**  
**Ad alte masse il picco monoisotopico può non essere quello più abbondante**



## Relative Isotope Abundance of Common Elements:

Isotope	Mass	[%] Abundance
1-H	1.007825	99.985
2-H (Deuterium)	2.014000	0.015
12-C	12.00000	98.90
13-C	13.00336	1.10
14-N	14.00307	99.63
15-N	15.00011	0.37
16-O	15.99491	99.76
18-O	17.99916	0.20
19-F	18.99840	100
23-Na	22.98977	100
31-P	30.97376	100
32-S	31.97207	95.03
34-S	33.96787	4.22
35-Cl	34.96885	76.77
37-Cl	36.96590	31.98
39-K	38.96371	93.26
79-Br	78.91834	50.69
81-Br	80.91629	49.31

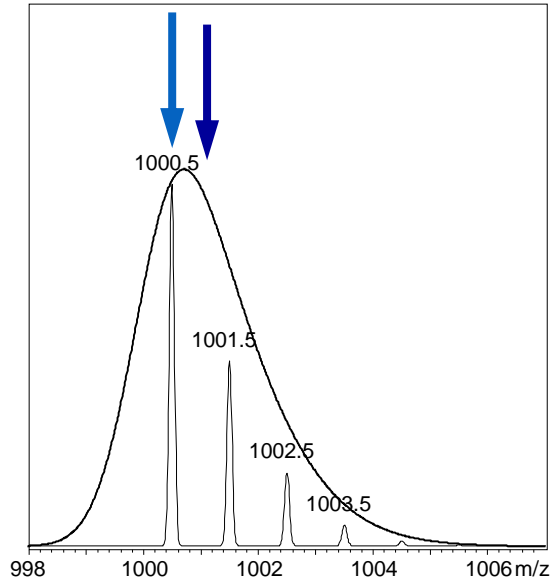
The monoisotopic mass is the mass of the isotopic peak whose elemental composition is composed of the most abundant isotopes of those elements

# MALDI-TOF

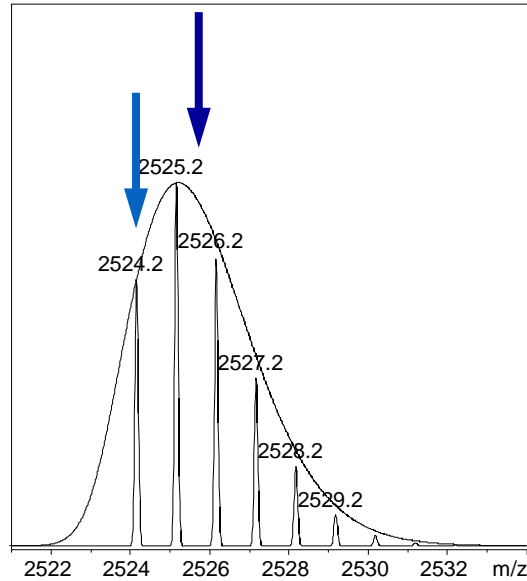
## Where do these **isotope peaks** originate from?

For a given molecule, the individual isotopes of all the elements contained in it finally yield a characteristic intensity distribution of isotopic masses. This is shown below by means of the isotopically resolved mass spectra of 3 compounds being different in size:

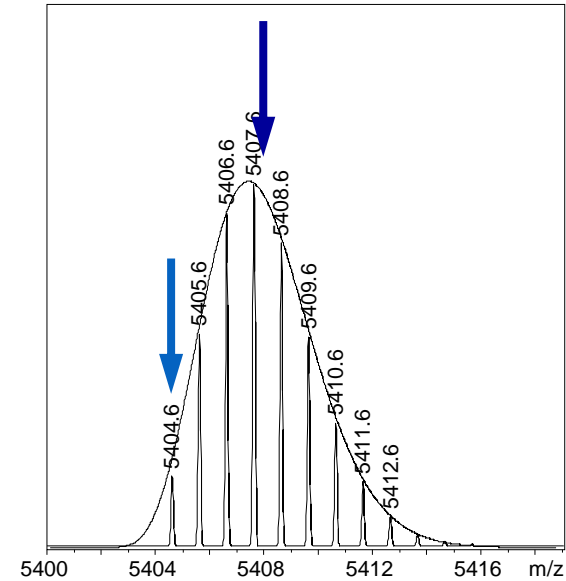
Element composition:  $C_{41}H_{69}N_{13}O_{14}S$   
 Monoisotopic mass  $[M+H]^+$ : 1000.4880  
 Average mass  $[M+H]^+$ : 1001.1409



Element composition:  $C_{112}H_{164}N_{29}O_{34}S_2$   
 Monoisotopic mass  $[M+H]^+$ : 2524.1510  
 Average mass  $[M+H]^+$ : 2525.8196



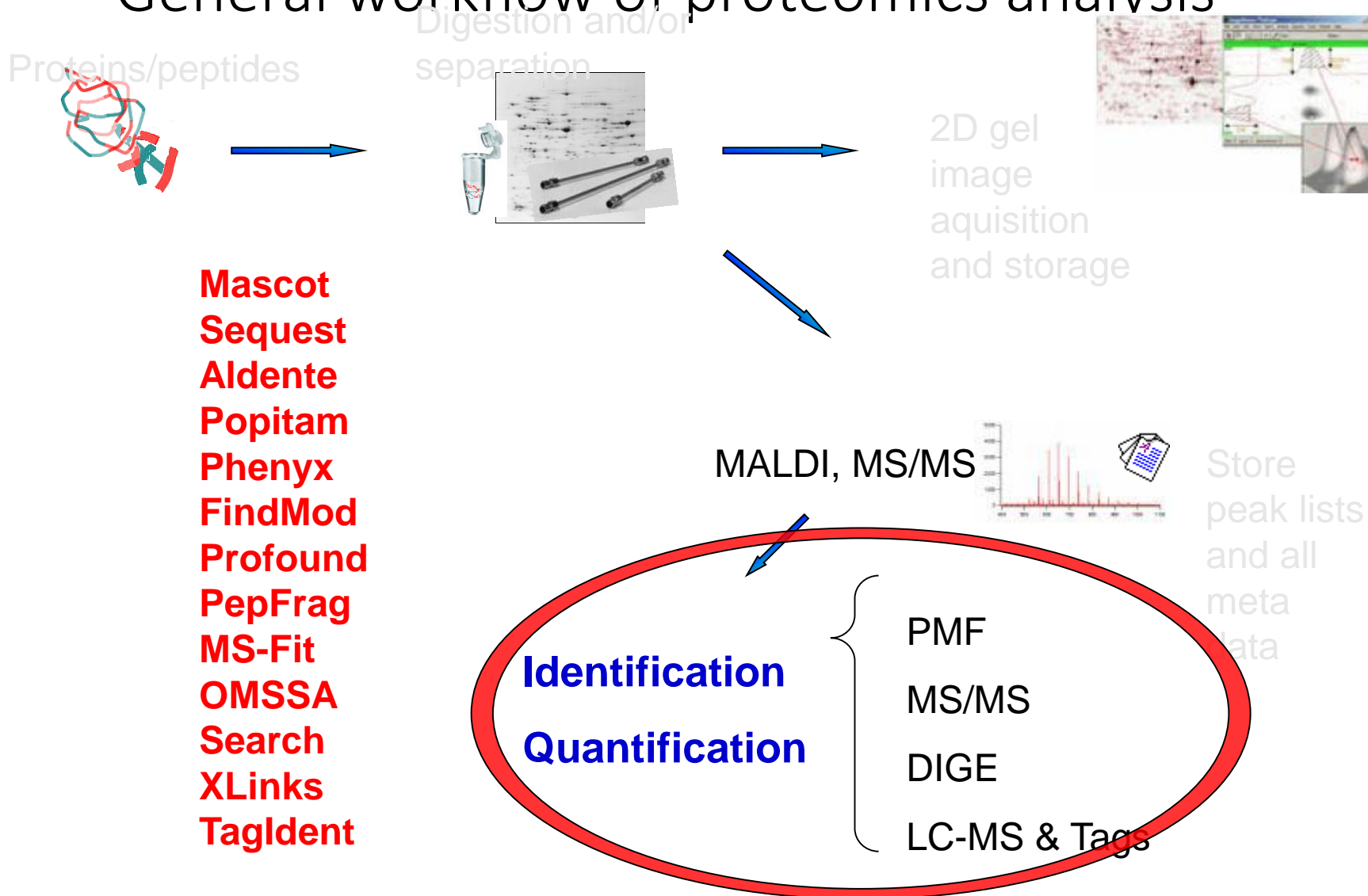
Element composition:  $C_{253}H_{363}N_{55}O_{75}S$   
 Monoisotopic mass  $[M+H]^+$ : 5404.6075  
 Average mass  $[M+H]^+$ : 5407.9984



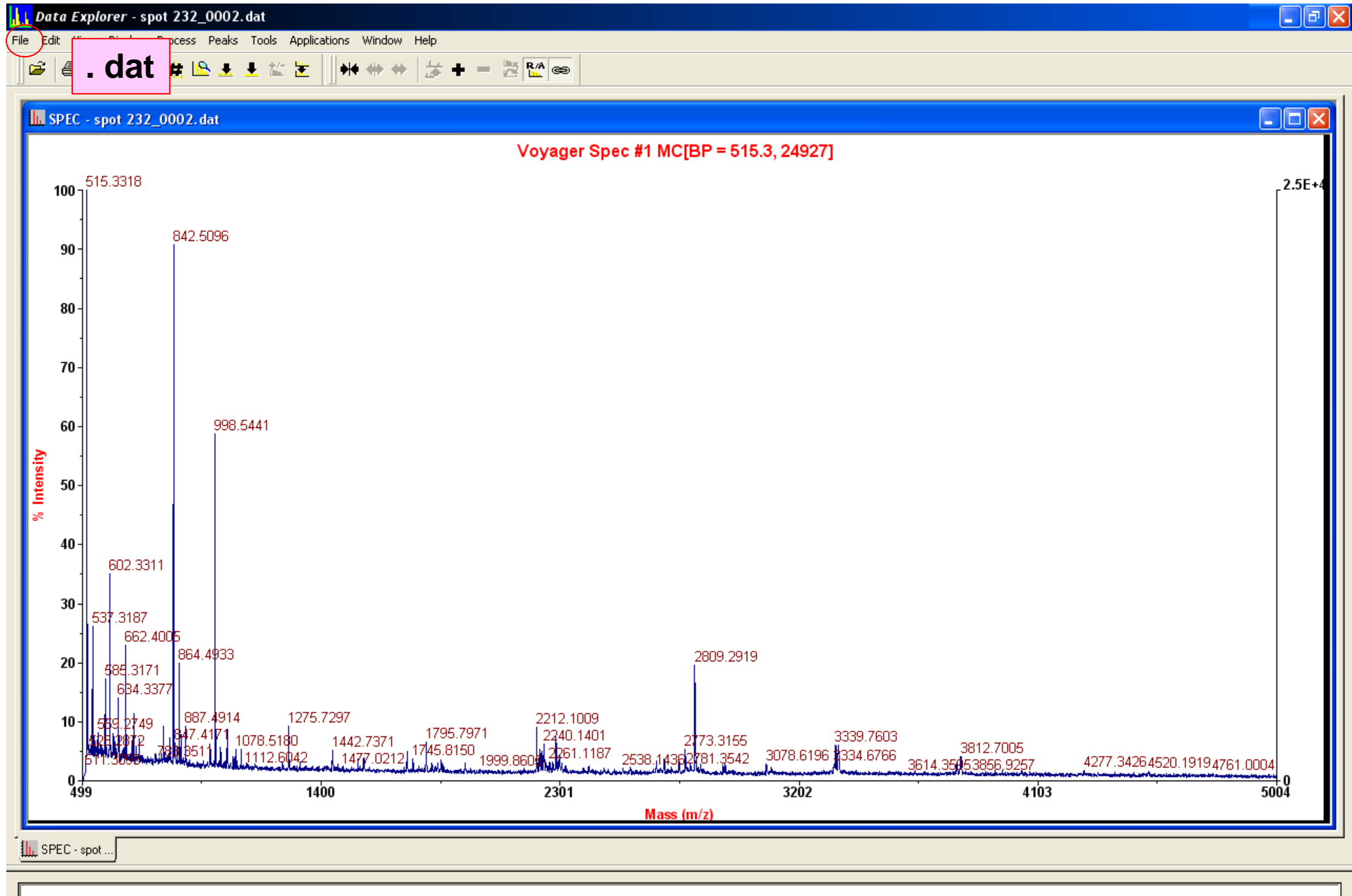
The **monoisotopic mass** is the sum of the masses of all the atoms present in a molecule using the mass of the most abundant isotope for each element.

The **average mass** of a molecule is the sum of elemental masses using the average weighted over all stable isotopes of each element contained in the molecule.

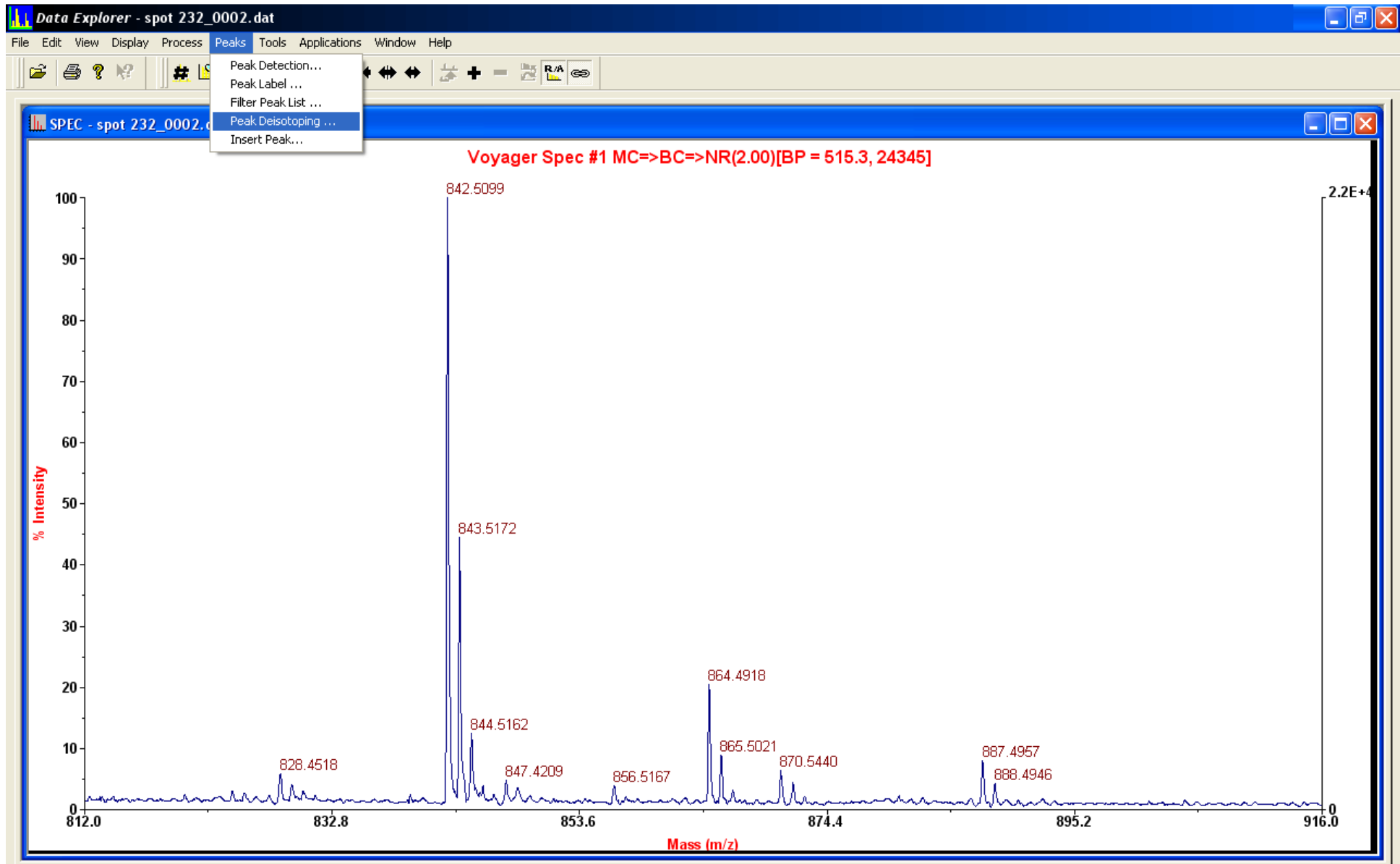
# General workflow of proteomics analysis

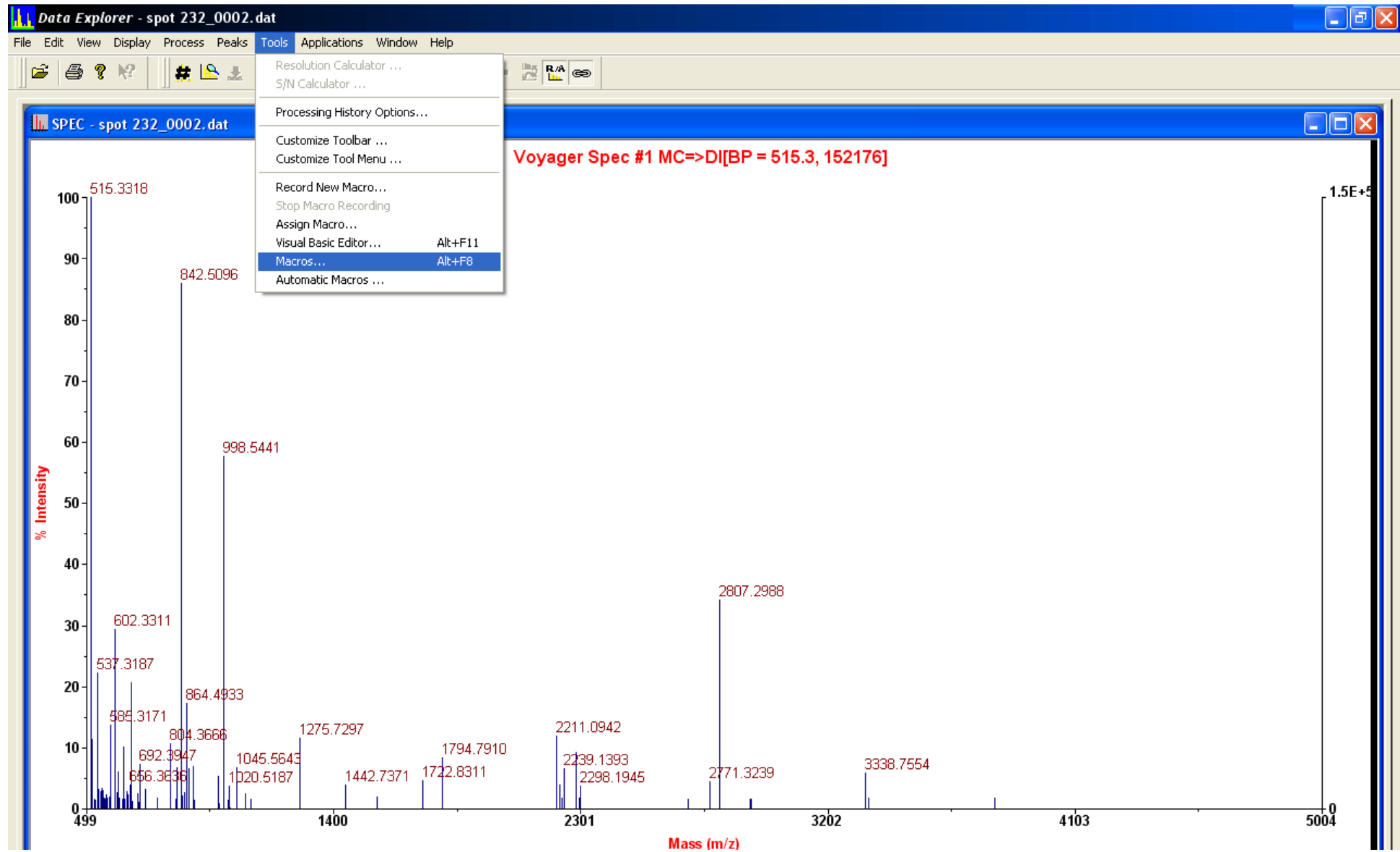


# Analisi dello spettro: Software DATA EXPLORER











# MASCOT Sequence Query

<b>Your name</b>	<input type="text"/>	<b>Email</b>	<input type="text"/>
<b>Search title</b>	<input type="text" value="spot 232_0002.dat - SpecView"/>		
<b>Database(s)</b>	Vertebrates_EST <input type="button" value="▲"/> Rodents_EST <input type="button" value="☰"/> Prokaryotes_EST Plants_EST Mus_EST <input type="button" value="▼"/>	<b>Enzyme</b>	Trypsin <input type="button" value="▼"/>
		<b>Allow up to</b>	1 <input type="button" value="▼"/> missed cleavages
		<b>Quantitation</b>	None <input type="button" value="▼"/>
<b>Taxonomy</b>	All entries <input type="button" value="▼"/>		
<b>Fixed modifications</b>	<input type="text" value="--- none selected ---"/> <input type="checkbox"/> Display all modifications	<input type="button" value="&gt;"/> <input type="button" value="&lt;"/>	<input type="button" value="▲"/> Acetyl (K) Acetyl (N-term) Acetyl (Protein N-term) Amidated (C-term) Amidated (Protein C-term) Ammonia-loss (N-term C) Biotin (K) Biotin (N-term) Carbamidomethyl (C) Carbamyl (K) Carbamyl (N-term) <input type="button" value="▼"/>
<b>Variable modifications</b>	<input type="text" value="--- none selected ---"/>	<input type="button" value="&gt;"/> <input type="button" value="&lt;"/>	<input type="button" value="▲"/> Biotin (K) Biotin (N-term) Carbamidomethyl (C) Carbamyl (K) Carbamyl (N-term) <input type="button" value="▼"/>
<b>Peptide tol. ±</b>	100 <input type="button" value="ppm ▼"/> # <sup>13</sup> C 0 <input type="button" value="▼"/>	<b>MS/MS tol. ±</b>	0.25 <input type="button" value="Da ▼"/>
<b>Peptide charge</b>	7- <input type="button" value="▼"/>	<b>Monoisotopic</b>	<input checked="" type="radio"/> Average <input type="radio"/>
<b>Query</b>	<input type="button" value="▲"/> SEARCH=PMF REPTYPE=Protein 514.29373 515.33184 517.323 537.3187 <input type="button" value="▼"/>		
<b>Instrument</b>	Default <input type="button" value="▼"/>		
<b>Decoy</b>	<input type="checkbox"/>	<b>Report top</b>	AUTO <input type="button" value="▼"/> hits
	<input type="button" value="Start Search ..."/>		<input type="button" value="Reset Form"/>

# MASCOT Sequence Query

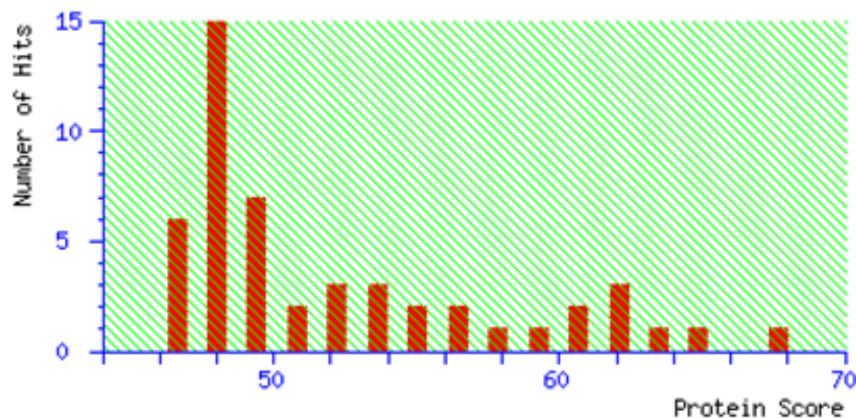
<b>Your name</b>	<input type="text" value="Patrizia"/>	<b>Email</b>	<input type="text" value="patriziacancemi@tin.it"/>
<b>Search title</b>	<input type="text" value="spot 232_0002.dat - SpecView"/>		
<b>Database(s)</b>	<input type="text" value="Human_EST"/> <input type="text" value="Fungi_EST"/> <input type="text" value="Environmental_EST"/> <input type="text" value="SwissProt"/> <input type="text" value="NCBIInr"/>	<b>Enzyme</b>	<input type="text" value="Trypsin"/>
		<b>Allow up to</b>	<input type="text" value="1"/> missed cleavages
		<b>Quantitation</b>	<input type="text" value="None"/>
<b>Taxonomy</b>	<input type="text" value="..... Homo sapiens (human)"/>		
<b>Fixed modifications</b>	<input type="text" value="Carbamidomethyl (C)"/>	<input type="button" value="&gt;"/> <input type="button" value="&lt;"/>	<input type="text" value="mTRAQ (N-term)"/> <input type="text" value="mTRAQ (Y)"/> <input type="text" value="mTRAQ:13C(3)15N(1) (K)"/> <input type="text" value="mTRAQ:13C(3)15N(1) (N-term)"/> <input type="text" value="mTRAQ:13C(3)15N(1) (Y)"/> <input type="text" value="NIPCAM (C)"/> <input type="text" value="Oxidation (HW)"/> <input type="text" value="Phospho (ST)"/> <input type="text" value="Phospho (Y)"/> <input type="text" value="Pyridylethyl (C)"/> <input type="text" value="Pyro-carbamidomethyl (N-term C)"/>
	Display all modifications <input type="checkbox"/>		
<b>Variable modifications</b>	<input type="text" value="Oxidation (M)"/> <input type="text" value="Propionamide (C)"/>	<input type="button" value="&gt;"/> <input type="button" value="&lt;"/>	
<b>Peptide tol. ±</b>	<input type="text" value="100"/> <input type="text" value="ppm"/>	<b># <sup>13</sup>C</b>	<input type="text" value="0"/>
		<b>MS/MS tol. ±</b>	<input type="text" value="0.25"/> <input type="text" value="Da"/>
<b>Peptide charge</b>	<input type="text" value="1+"/>	<b>Monoisotopic</b>	<input checked="" type="radio"/> Average <input type="radio"/>
<b>Query</b>	<input type="text" value="SEARCH=PMF"/> <input type="text" value="REPTYPE=Protein"/> <input type="text" value="514.29373"/> <input type="text" value="515.33184"/> <input type="text" value="517.323"/> <input type="text" value="537.3187"/>		
<b>Instrument</b>	<input type="text" value="Default"/>		
<b>Decoy</b>	<input type="checkbox"/>	<b>Report top</b>	<input type="text" value="AUTO"/> hits
<input type="button" value="Start Search ..."/>		<input type="button" value="Reset Form"/>	

# *MATRIX* SCIENCE Mascot Search Results

User : Patrizia  
 Email : patriziacancemi@tin.it  
 Search title : spot 232\_0002.dat - SpecView  
 Database : SwissProt 2010\_05 (516603 sequences; 181919312 residues)  
 Timestamp : 14 May 2010 at 16:26:32 GMT  
 Top Score : 68 for **CAN3\_MACFA**, Calpain-3 OS=Macaca fascicularis GN=CAPN3 PE=2 SV=1

## Mascot Score Histogram

Protein score is  $-10 \cdot \log(P)$ , where P is the probability that the observed match is a random event.  
 Protein scores greater than 70 are significant ( $p < 0.05$ ).



## Protein Summary Report

Format As	Protein Summary <input type="button" value="v"/>	<a href="#">Help</a>
Significance threshold p<	<input type="text" value="0.05"/>	Max. number of hits <input type="text" value="AUTO"/>

513.28032	601.33367	1020.52183
514.31182	602.3327	1045.56489
515.33208	603.34387	1078.52167
517.32344	605.33099	1097.583
519.34464	607.30016	1254.68919
520.30616	610.32244	1275.73388
521.31843	613.35035	1341.71444
524.27141	616.37186	1442.72853
525.28402	617.37586	1537.65441
527.25242	618.32429	1559.66478
528.3148	623.37918	1722.82836
529.26915	624.31278	1744.85151
530.3145	625.36789	1794.79203
531.30499	629.26353	1816.84111
532.29962	630.36604	1829.95287
536.28928	633.29828	1837.874
537.31947	634.33652	1851.83742
539.02839	635.34737	1940.95497
540.27901	646.37683	2211.1046
541.29915	652.32833	2221.12513
545.00138	653.35786	2222.11765
545.26798	655.32927	2225.156
547.26952	656.37218	2233.0661
548.28543	657.37053	2239.13929
549.2304	661.37621	2261.13234
550.32061	662.40108	2283.19558
551.28686	683.39374	2284.17151
552.30398	684.36785	2292.15585
553.29473	687.38667	2297.20997
558.24968	691.39037	2305.15757
559.32232	692.39785	2408.17397
560.31588	701.48289	2691.27519
563.35092	707.4208	2748.35138
565.33119	714.38991	2751.18472
567.0281	723.33645	2771.39975
568.20009	772.4391	2791.37256
569.30876	804.36597	2795.13281
575.33414	805.38676	2807.31599
579.3513	806.40291	2921.33481
581.32657	824.43486	3096.8624
582.27999	827.50568	3333.61437
584.33389	828.45179	3338.79755
585.31873	842.5099	3348.5959
586.33678	846.3923	3352.44528
589.21695	847.42092	3795.0807
590.24667	856.5167	3795.9656
591.32277	864.49177	3809.8987
592.31411	870.54404	3811.76443
596.26551	886.46612	3857.26716
598.30333	887.49573	4274.90814
599.17752	892.49859	4521.35947
600.33234	980.54747	4990.4464
	998.54434	

## Selezione manuale dei picchi di interesse

Addotti del Na = + 22

Addotti del K = + 39

Addotti dell'O = + 16

Masse  $\leq$  a 600 = frammenti dimatrice

Masse  $\geq$  2280 = polimeri di matrice

**Regola empirica:**

Peptidi con 3 cifre = il primo decimale deve essere la metà del primo valore

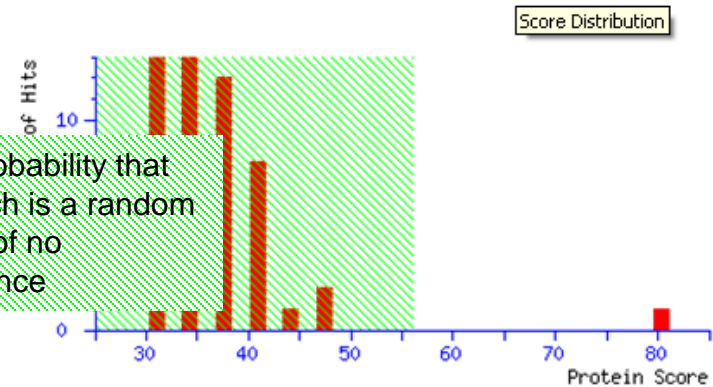
Peptidi con 4 cifre = il primo decimale deve essere la metà della dei primi due valori

# *MATRIX* SCIENCE Mascot Search Results

User : Patrizia  
 Email : patriziacancemi@tin.it  
 Search title : spot 232\_0002.dat - SpecView  
 Database : SwissProt 2010\_05 (516603 sequences; 181919312 residues)  
 Taxonomy : Homo sapiens (human) (20278 sequences)  
 Timestamp : 14 May 2010 at 16:24:12 GMT  
 Top Score : 80 for **GELS\_HUMAN**, Gelsolin OS=Homo sapiens GN=GSN PE=1 SV=1

## Mascot Score Histogram

Protein score is  $-10 \cdot \log(P)$ , where P is the probability that the observed match is a random event.  
 Protein scores greater than 56 are significant ( $p < 0.05$ ).



> 5% probability that the match is a random events, of no significance

The significance of the result depends on the size of the database being searched

## Protein Summary Report

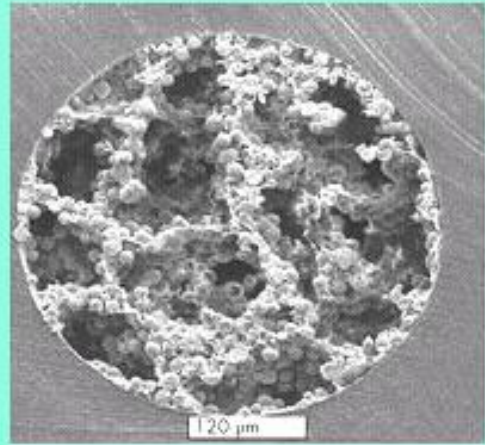
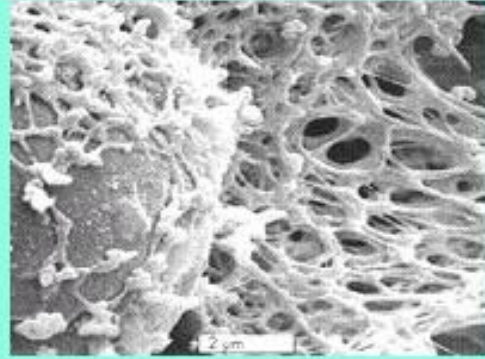
Format As: Protein Summary

Significance threshold p<  Max. number of hits

## Index

	Accession	Mass	Score	Description
1.	<a href="#">GELS_HUMAN</a>	86043	80	Gelsolin OS=Homo sapiens GN=GSN PE=1 SV=1
2.	<a href="#">CHD1L_HUMAN</a>	101507	48	Chromodomain-helicase-DNA-binding protein 1-like OS=Homo sapiens GN=CHD1L PE=1 SV=1

## Zip-Tip



Per purificare e concentrare in pochi secondi da femtomoli a picomoli di campioni biologici ZipTip è un puntale per pipetta da 10  $\mu\text{L}$  con un letto di resina cromatografica di circa 0,6  $\mu\text{L}$ , fissato in punta senza volumi morti. E' ideale per concentrare, dissalare e frazionare in pochi secondi peptidi, proteine od oligonucleotidi da analizzare in spettrometria di massa, HPLC, elettroforesi capillare ed altre tecniche analitiche. ZipTip è semplice e facile da usare.

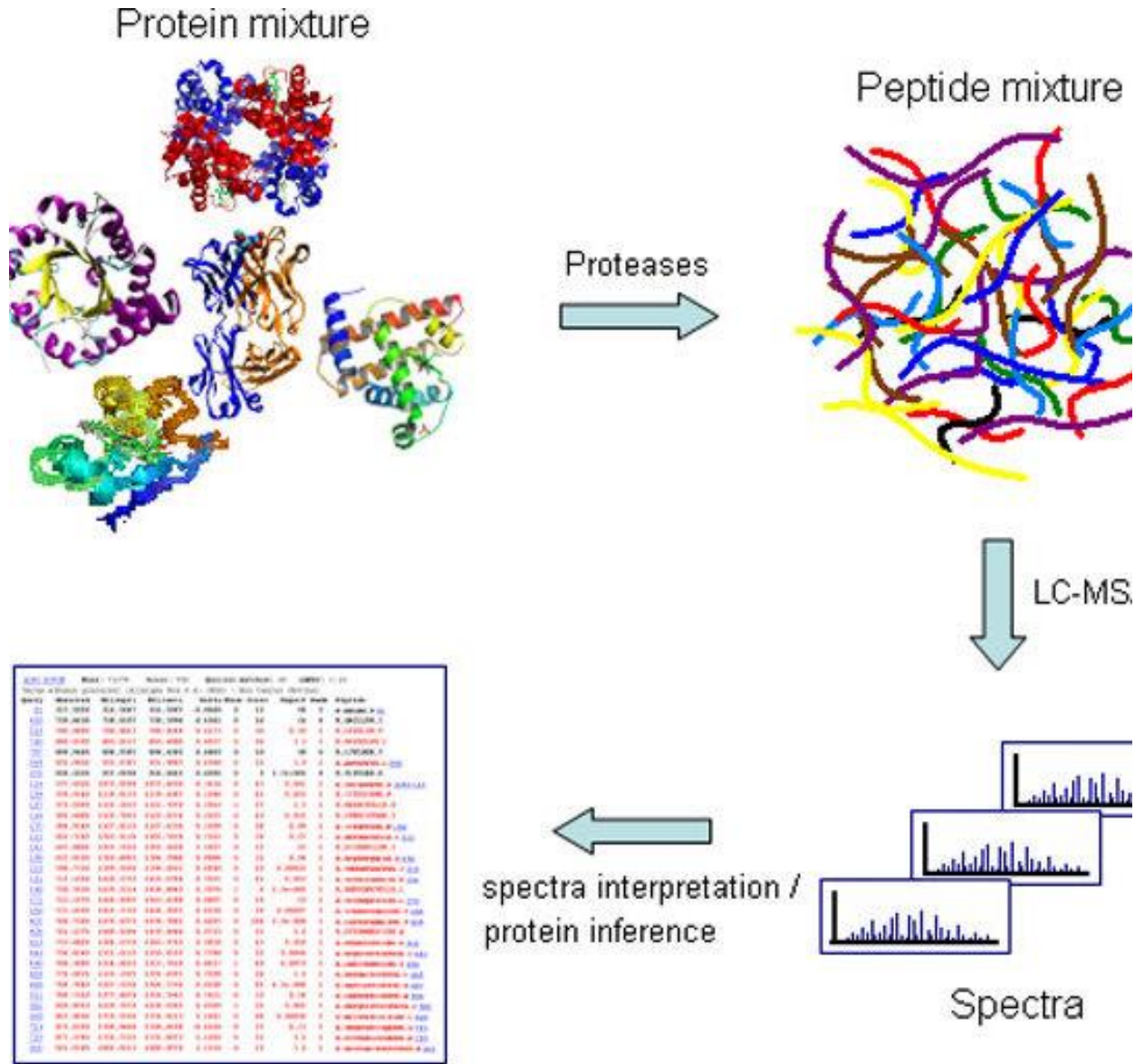
Dopo averlo saldamente alloggiato in punta ad una pipetta singola o multicanale, ad un ago cilindrico calibro 22 di siringa HPLC oppure ai coni di un sistema automatizzato, per legare il campione occorre aspirare ed erogare il liquido diverse volte attraverso la resina. Allo stesso modo, è possibile dilavare i contaminanti e poi eluire con precisione il campione purificato in 1 - 4  $\mu\text{L}$  di solvente compatibile con il trasferimento diretto sul target dello spettrometro di massa o in provettina.

Quando c'è il bisogno d'eluire in volumi ancora inferiori (es.:  $<1 \mu\text{L}$ ), si può ricorrere allo ZipTip con micro-letto di soli 0,2  $\mu\text{L}$  di resina. ZipTip mette a disposizione riproducibilità ed alta resa per purificare e concentrare pico e femtomoli di peptidi, proteine od oligonucleotidi e quindi migliorare la qualità del dato analitico; con la possibilità di effettuare eluizioni a gradiente, è anche possibile, per esempio, semplificare l'analisi di miscele complesse di peptidi frazionandole. ZipTip è un dispositivo veloce e sempre pronto all'uso che fa risparmiare tempo sostituendo, la classica tecnica cromatografica, in tutte le preparazioni del campione tranne quelle più complesse.

# La proteomica gel-free o shot gun proteomics

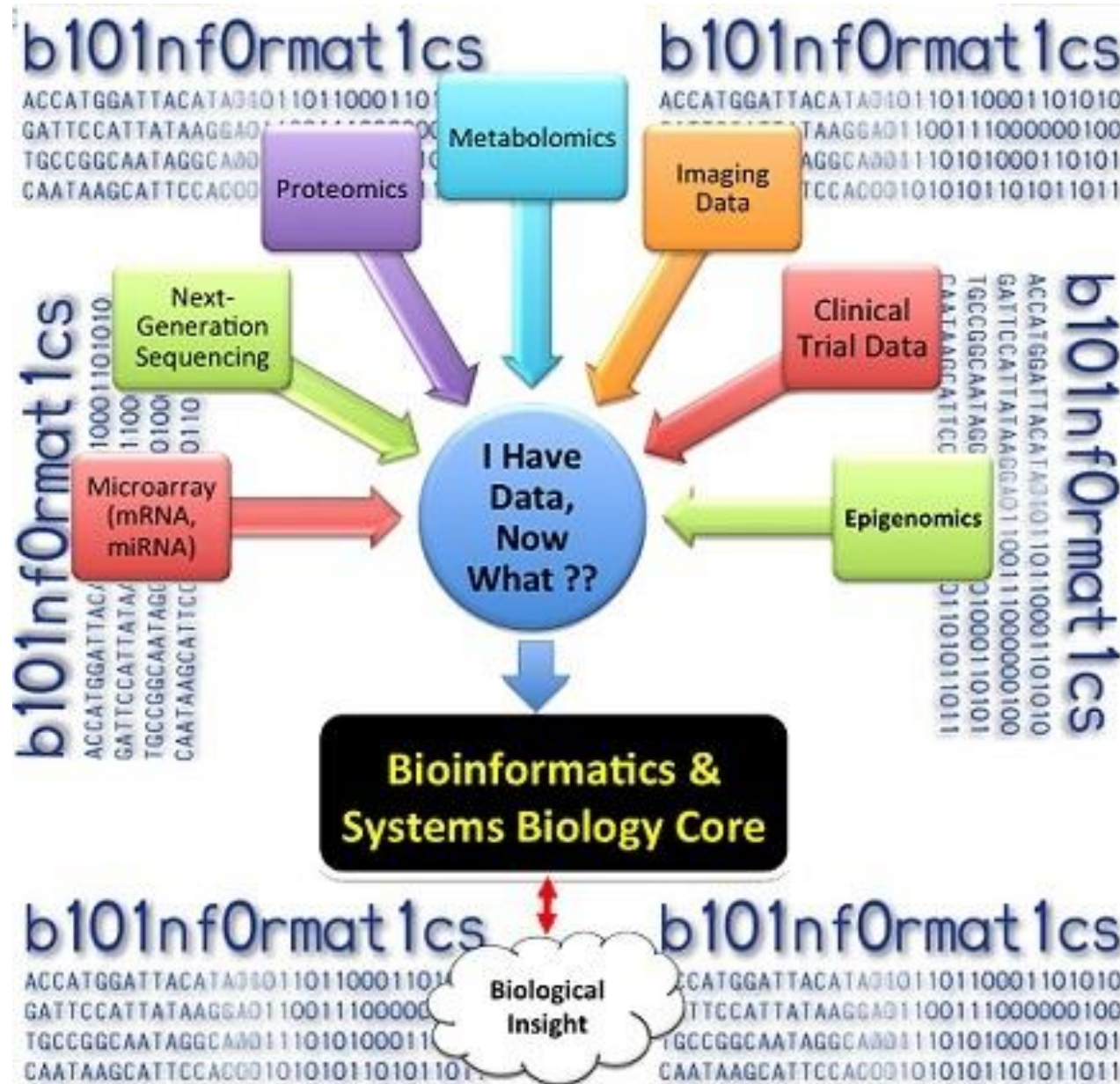
I dati omici che derivano dalla proteomica shot gun sono molto complessi, in quanto contemporaneamente abbiamo informazioni di espressione e modulazione di centinaia di proteine

CLUSTERIZZAZIONE DELLE PROTEINE  
IN BASE ALLA FUNZIONE



tein identification / quantitation

# System biology



La *systems biology*, quindi, ha come obiettivo lo studio dei meccanismi attraverso cui le macromolecole cellulari interagiscono in modo dinamico fra loro per generare le proprietà funzionali delle cellule viventi

- La complessità strutturale delle cellule può essere disassemblata in moduli, collegati da connettori modulari
- costruzione di modelli di moduli singoli che poi verranno assemblati per dare via via strutture più complesse.



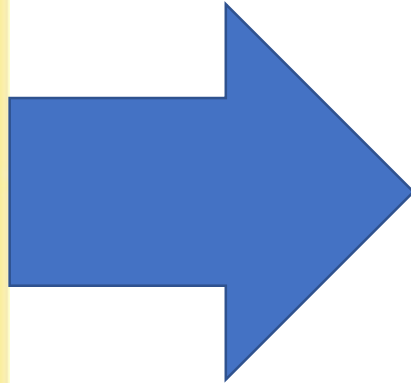
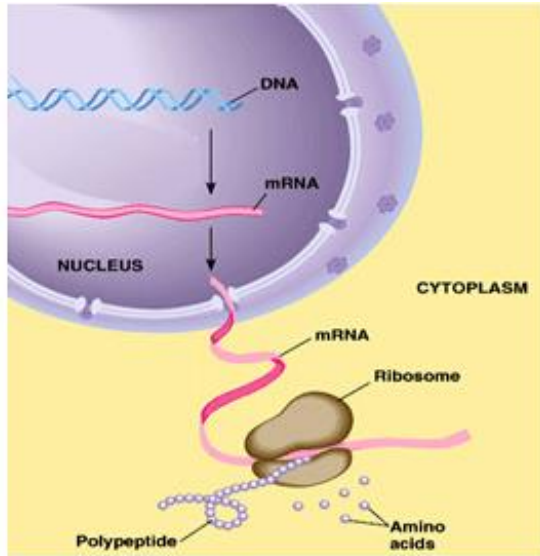


# L'approccio con lo studente

---



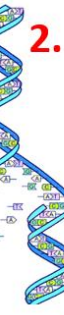
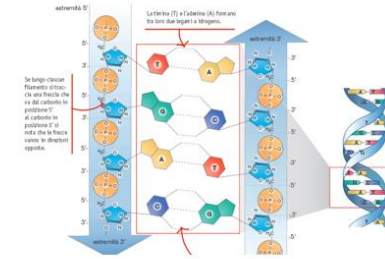
# Dai geni alle proteine: il ruolo della bioinformatica



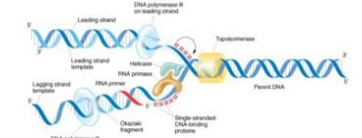
UNIVERSITÀ  
DEGLI STUDI  
DI PALERMO

Scienze omiche: Grandi apparecchiature per la Proteomica

## 1. Struttura chimica del DNA

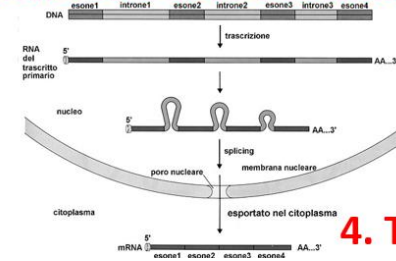


## 2. Replicazione del DNA



Il DNA si replica grazie ad un modello definito semiconservativo

## 3. Trascrizione del DNA in RNA




## 4. Traduzione dell'mRNA in proteine

		SECONDA "LETTERA"					
		U	C	A	G		
U	UUU	PHE	UCU	UAU	TYR	UGU	CYS
	UUC	UCG	UAC	UGC	UGA	STOP	A
	UUA	UCA	UAA	STOP	UGG	TRP	G
	UUG	UCG	UAG	STOP	UGG	TRP	G
C	CUU	CCU	CAU	HIS	CGU	ARG	A
	CUC	CCC	CAC	GLN	CGC	ARG	A
	CUA	CCA	CAA	GLN	CGG	ARG	A
	CUG	CCG	CAG	GLN	CGG	ARG	A
A	AUU	ACU	AAU	ASN	AGU	SER	U
	AUC	ACC	AAC	ASN	AGC	SER	U
	AUA	ACA	AAA	LYS	AGA	ARG	A
	AUG	ACG	AAG	LYS	AGG	ARG	A
G	GUU	GCU	GAU	ASP	GGU	GLY	A
	GUC	GCC	GAA	ASP	GGC	GLY	A
	GUA	GCA	GAA	ASP	GGG	GLY	A
	GUU	GCC	GAA	ASP	GGG	GLY	A

Le regole del codice genetico, per tradurre i codoni di RNA in aminoacidi.

## BANCHE DATI

Consentire la consultazione e l'analisi delle informazioni in essa contenute e di ogni altra informazione a esse correlate e memorizzate in altre banche dati


National Library of Medicine  
National Center for Biotechnology Information
Log in

All Databases ▾

- All Databases
- Assembly
- Biocollections
- BioProject
- BioSample
- Books
- ClinVar
- Conserved Domains
- dbGaP
- dbVar
- Gene
- Genome
- GEO DataSets
- GEO Profiles
- GTR
- HomoloGene
- Identical Protein Groups
- MedGen
- MeSH
- NLM Catalog

- MeSH
- NLM Catalog
- Nucleotide
- OMIM
- PMC
- PopSet
- Protein
- Protein Clusters
- Protein Family Models
- PubChem BioAssay
- PubChem Compound
- PubChem Substance
- PubMed
- SNP
- SRA
- Structure
- Taxonomy
- ToolKit
- ToolKitAll
- ToolKitBookgh

Search


**NCBI Home**

**Resource List (A-Z)**

- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation


**Download**

Download NCBI data to your computer




**Learn**

Find help documents, attend a class or watch a tutorial




**Research**

Explore NCBI research and collaborative projects




**Develop**

Use NCBI APIs and code libraries to build applications



**Analyze**

Identify an NCBI tool for your data analysis task



**Popular Resources**

- [PubMed](#)
- [Bookshelf](#)
- [PubMed Central](#)
- [BLAST](#)
- [Nucleotide](#)
- [Genome](#)
- [SNP](#)
- [Gene](#)
- [Protein](#)
- [PubChem](#)

---

**NCBI News & Blog**

- Announcing the NCBI Datasets SARS-CoV-2 taxonomy page 26 Jul 2022
- Need SARS-CoV-2 assembled genome 18 Jul 2022
- RefSeq Release 213 18 Jul 2022

RefSeq Release 213 is now available online, from the FTP site and through

Try out the latest BLAST ClusteredNR database results. Now with in-cluster

## Basic Local Alignment Search Tool

**BLAST** finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

NEWS

### BLAST+ 2.13.0 is here!

Starting with this release, we are including the blastn\_vdb and tblastn\_vdb executables in the BLAST+ distribution.

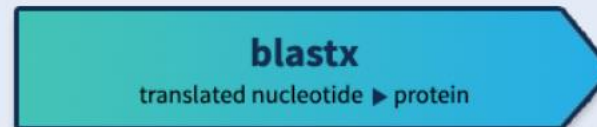
Thu, 17 Mar 2022 12:00:00 EST

[More BLAST news...](#)

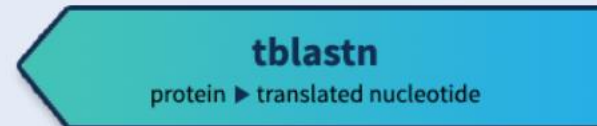
## Web BLAST



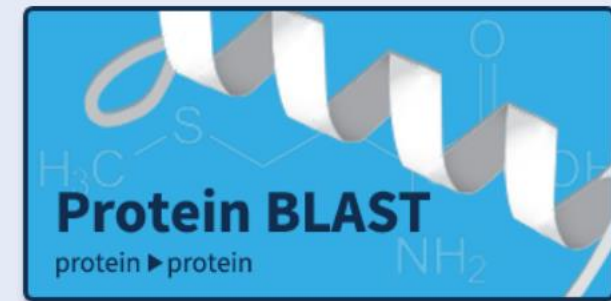
**Nucleotide BLAST**  
nucleotide ► nucleotide



**blastx**  
translated nucleotide ► protein



**tblastn**  
protein ► translated nucleotide



**Protein BLAST**  
protein ► protein

# Find your protein

UniProtKB ▾

[Advanced](#) | [List](#) [Search](#)

Examples: [Insulin](#), [APP](#), [Human](#), [P05067](#), [organism\\_id:9606](#)

UniProt is the world's leading high-quality, comprehensive and freely accessible resource of protein sequence and functional information. [Cite UniProt](#)”



Accessing UniProt programmatically? Have a look at the [new API documentation](#).

If you still need it, the [legacy version of the website](#) is available until the 2022\_03 release.

[Proteins](#)

[Species](#)

[Protein Clusters](#)

[Sequence Archive](#)